الدكتور عبد الحميد عبد المجيد البلداوي

الأساليب التطبيقية لتحليل وإعداد البحوث العلمية

مع حالات دراسية باستخدام برنامج SPSS

2014 dgh

DIGITAL SCIENTIFIC RESEARCH

dtg 2000d

155400 lkop

الشروق

المحتـــويات

مقـدمـة Introduction 21

الفصل الأول: تمهيد ومفاهيم اساسية 23

Preface and Basic Definitions

1-1 أهمية البحوث العلمية 23

2-1 مفهوم البحث العلمي وخصائصه 24

Research Definition and Properties Scientific..........

1-2-1 الموضوعية Objectivity 24

1-2-2 الاستناد للحقائق Facts 24

1-2-3 التعميم Generalization 25

1-2-4 التوافق بين الاستنباط والاستقراء

Induction & Deduction Matching 25

1-2-5 القدرة على بناء التنبوءات Predictive Ability 25

1-2-6 تنوع سبل إنجاز البحوث 25

(1) البحث النظري Theoretical Research 25

(2) البحث التطبيقي Applied Research 26

(3) بحث مزيج بين النظرية والتطبيق 26

3-1 أنواع المعطيات الاحصائية 26

1-3-1 المعطيات الكمية Quantitative Data 26

(1) معطيات كمية متصلة 26

continuous Quantitative variables ....

(2) معطيات كمية متقطعة 27

discrete Quantitative variables ....

1-3-2 المعطيات النوعية Qualitative Data 28

(1) المتغيرات الاسمية Nominal Variables.................... 28
(2) المتغيرات الترتيبية Ordinal Variables...................... 29
1-4 مفهوم النماذج الاحصائية واصنافها 30
Statistical Models Definition and Classification.........
1-4-1 تصنيف حسب نوع المعطيات المستخدمة.................... 30
1-4-2 تصنيف حسب صيغة المشاهدات................... 31
1-4-3 تصنيف وفقا للهدف من البحث.................... 31
1-4-4 تصنيف حسب نوع المعادلة المستخدمة 31
1-5 الاهداف العامة للبحوث General Goals of Research 31
1-5-1 الوصف والتفسير Description and Explanation....... 31
1-5-2 بناء تقديرات وتوقعات Forecasting &Estimation..... 34
1-5-3 السيطرة والتحكم Controlling & Judgmen 34
1-5-4 اختبار الفروض Hypotheses Testing................... 36
1-6 منهجيات البحث العلمي
Scietific Research Methodologies............................ 36
1-6-1 منهجية المفردة Disaggregate Methodology............ 36
(1) المعطيات السلوكية Behavioural Data.................... 36
(2) معطيات الرأي او الاعتقاد Attitudinal Data.............. 37
(3) حالة دراسية رقم C_{1-1}: استخدام برنامج SPSS في تحويل المعطيات الغير رقمية الى رقمية 37
1-6-2 المنهجية التجميعية Aggregate Methodology 48
1-6-3 حالة دراسية رقم C1-2: استخدام برنامج SPSS في الحصول على معطيات على مستوى المفردة والمستوى التجميعي 51
1-7 صياغة المتغيرات Variables Formulation................... 54
1-7-1 الصيغة المطلقة Absolute Form........................... 55
1-7-2 الصيغة النسبية Ratio Form 55

1-7-3 صيغة الفرق المطلق Absolute Difference Form 55
1-7-4 صيغة الفرق المطلق المرجح
Weighted Absolute Difference Form 55
1-7-5 صيغ اخرى 55
1-8 أنواع الاستقصاءات الاحصائية 55
1-8-1 المسوحات الشاملة والمعلمة Census & Parameter 55
1-8-2 مسوحات العينة والاحصاءه
Sampling Surveys & Statistic 55
1-9 اطار المجتمع الاحصائي
Statistical Population Framework 59

الفصل الثاني

تصميم الاستبانة الاحصائية و طرق جمع المعطيات

Questionnaire Design & Data Collection Methods

61
2-1 التغطية او الشمولية Coverage 61
2-1-1 التغطية المكانية Spatial Coverage 61
2-1-2 التغطية النوعية Data Coverage 61
2-2 القواعد العامة لتصميم الاستبانة
General Rules of Questionnaire Design 62
2-2-1 ان يكون حجمها مناسبا 62
2-2-2 مراعاة التنفيذ الالي لتبويب المعطيات 62
2-2-3 ان تضم الحد الامثل من الاسئلة 62
2-2-4 مراعاة المفاهيم والتصانيف الاحصائية الدولية 63
2-3 شروط صياغة اسئلة الاستبانة 63
2-4 الاجزاء التي تتكون منها الاستبانة 66
2-5 حالات دراسية في تصميم الاستبانة 67

2-5-1 حالة دراسية C_{2-1}: في تصميم أستبانة تتكون من ثلاثة اجزاء... 67
5-5-2 حالة دراسية C_{2-2} : في تصميم أستبانة تتكون من جزئين...... 71
2-6 طرق جمع المعطيات
Methods of Data Collection 75
2-5-1 طريقة المشاهدة Observation Method 75
2-5-2 طريقة التسجيل الذاتي Self-Recording Method 75
2-5-3 طريقة المقابلة الشخصية Interviewing Method. 77
2-5-4 طريقة الهاتف Telephone Method 77
2-5-5 طريقة التركيز على المناقشات الجماعية
Focus Group Discussion Method 78

الفصل الثالث

تصميم العينات واسلوب تحديد حجمها 79

Sampling Design and Sample Size Determination

3-1 مفهوم تصميم العينة 79
3-2 مفهوم العينات العشوائية Random Samples.................... 80
3-3 العينة العشوائية البسيطة Simple Random Sample 81
3-3-1 مفهوم العينة العشوائية البسيطة......... 81
3-3-2 حالات استخدامها.................. 82
3-3-3 اساليب اختيار العينة Sample Selection Methods 82
3-3-4 اساليب الاختيار العشوائي لوحدات العينة
Methods of Random Selection of Observations 82
3-3-5 عيوب العينة وميزاتها......................... 86
3-4 تقدير معالم المجتمع من نتائج العينة العشوائية البسيطة
Population Parameters Estimation 87
3-4-1 أنواع القيم التقديرية................. 87

(1) التقدير بنقطة **Point Estimation** 88
(2) التقدير بفترة **Interval Estimation** 89
3-4-2 فترة الثقة لمتوسط المجتمع
Confidence Intervals of μ μ, ... 89
(1) فترة الثقة لمتوسط مجتمع معلوم التباين 89
(2) فترة الثقة لمتوسط مجتمع مجهول التباين 91
(3) فترة الثقة لنسبة خاصية مجتمع 93
3-4-3 فترة الثقة لتباين المجتمع
Confidence Intervals of σ^2 σ^2 94
3-4-4 حالة دراسية رقم $C_{3\text{-}1}$ في توزيع المعاينة
Samplin Distribution 97
3-5 تحديد حجم العينة **Sample Size Determination** 100
3-5-1 طريقة حساب حجم العينة
Method of Sample Size Calculation 101
3-5-2 طريقة حساب حجم العينة لحالة النسب 103
Sample size Calculation in Case of Percentage ....
(1) في حالة معلومية نسبة خاصية المجتمع وحجمه 103
(2) في حالة معلومية نسبة خاصية المجتمع ومجهولية حجمه 104
(3) في حالة مجهولية حجم المجتمع ونسبة خاصيته 104
3-5-3 طريقة حساب حجم العينة وفق الميزانية المتاحة
Sample size Calculation Aaccording to survey Budget 105
3-6 العينة العشوائية الطبقية **Stratified Random Sample** 106
3-6-1 مفهوم العينة حالات استخدامها 106
3-6-2 طريقة تحديد عدد وحدات العينة لكل طبقة 106
(1) طريقة الاختيار المتناسب 107
(2) طريقة الاختيار الامثل 109
3-6-3 تقدير متوسط وتباين المجتمع σ^2 and μ Estimations 112

(1) تقدير متوسط المجتمع..μ 112
(2) تقدير تباين المجتمع .. σ^2 112
3-7 العينة العشوائية المنتظمة Systematic Random Sample 112
3-7-1 مفهوم العينة وحالات استخدامها.. 112
3-7-2 اسلوب اختيار العينة Method of Sample Selection 113
3-7-3 الخطأ المعياري وتقدير مجموع لمجتمع
S.E. And N Estimations 115
3-8 العينة العشوائية العنقودية Cluster Random sample 115
3-8-1 مفهوم العينة وحالات استخدامها.. 115
3-8-2 اسلوب اختيار العينة Method of the Sample Selection 116
3-8-3 تقدير المتوسط وتباين المجتمع من نتائج العينة العنقودية
σ^2 and μ Estimations 116
3-9 العينة العشوائية المتعددة المراحل وحالات استخدامها
Multi-Stage Random Sample and cases of Use 119
3-9-1 مفهوم العينة واستخداماتها 119
3-9-2 تقدير متوسط المجتمع وتباينه 121
3-10 حالات دراسية في العينات العشوائية 122
3-10-1 حالة دراسية رقم C_{3-2} : في تصميم عينة لنشاط الخدمات
المالية 122
3-10-2 حالة دراسية رقم C_{3-3} : في تحديد حجم عينة مسحوبة من
جتمع مجهول المعالم.. 122
3-11 العينات غير العشوائية Non-Random Sample................ 128
3-11-1 مفهوم العينات و حالات استخدامها 128
3-11-2 انواع العينات غير العشوائية
Types of non-Random Samples 128
(1) العينة التحكمية (المتعمدة) 128
(2) العينة الحصصية.. 129

الفصل الرابع

131 تدقيق أخطاء نتائج الاستقصاء (المسح الاحصائي)

وطرق معالجة وتعويض المعطيات المفقودة

Survey Results Verifying and Missing Data Imputations

131 .... 1-4 تدقيق الاستبيانات **Questionnaires Verifying**

132 .... 2-4 تدقيق نتائج أخطاء المسح **Survey Errors Verifying**

132 .... 1-2-4 أخطاء المعاينة Sampling errors

133 .... 2-2-4 اخطاء غير المعاينة Non-Sampling Errors

136 .... 3-4 طرق تعويض المعطيات المفقودة **Imputation Methods**

1-3-4 طريقة التعويض الاستنتاجي

136 .... Deductive Imputation Method

2-3-4 التعويض باستخدام المعدل العام

137 .... Overall Mean Imputation Method

3-3-4 التعويض بمعدل فئة الرقم المفقود

137 .... Class Mean Imputation Method

4-3-4 التعويض العشوائي العام

138 .... Random Overall Imputation Method

5-3-4 التعويض العشوائي ضمن الفئة

138 .... Random imputation Within Class Method

6-3-4 طريقة المسافة التوفيقية

138 .... Distance Function Matching Method

138 .... 7-3-4 طريقة الانحدار Regression Analysis Method

4-4 استخدام النسب في تقدير مجموع ومتوسط المجتمع

140 .... **Ratios Use for Estimation**

143 .... 5-4 حالة دراسية رقم C_{4-1} في استخدام الانحدار في التقدير والتعويض

الفصل الخامس

التحليل الوصفي باستخدام مخرجات SPSS 151

Descriptive Analysis Using SPSS Output

5-1 مقاييس النزعة المركزية وغير المركزية Central and Non-Central Tendency Measures 151

5-1-1 المقاييس المركزية (المتوسطات) 151

5-1-2 المقاييس غير المركزية........................ 155

5-1-3 مقاييس التشتت................................ 157

5-1-4 خواص الانحراف المعياري.. 158

5-1-5 العرض البيان.. 160

5-1-6 حالة دراسية رقم C_{5-1} : استخدام برنامج SPSS في المقاييس الوصفية 165

5-2 تحليل الارتباط Correlation Analysis 172

5-2-1 الارتباط البسيط Simple Correlation 174

(1) حالات استخدام معامل الارتباط البسيط وصيغة حسابه........ 174

(2) حالة دراسية رقم C_{5-2} باستخدام برنامج SPSS 174

5-2-2 الارتباط المتعدد Multiple Correlation 177

(1) حالات استخدام معامل الارتباط المتعدد وصيغة حسابه........ 177

(2) حالة دراسية رقم C5-3 باستخدام برنامج SPSS............ 177

5-2-3 الارتباط الجزئي Partial Correlation...................... 181

(1) حالات استخدام معامل ارتباط الجزئي وصيغة حسابه 181

(2) حالة دراسية رقم C5-4 باستخدام برنامج SPSS 181

5-2-4 ارتباط الرتب Rank Correlation........................... 184

(1) حالات استخدام معامل ارتباط الرتب وصيغة حسابه 184

(2) حالة دراسية رقم C5-5 باستخدام برنامج SPSS. 184

5-2-5 ارتباط الاقتران Association Correlation 186

6-2-5 ارتباط التوافق Contingency Correlation 187

(1) حالات استخدام معامل ارتباط التوافق وصيغة حسابه 187

(2) حالة دراسية رقم C5-6 باستخدام برنامج SPSS 188

5-3 الانحدار في التحليل الوصفي

Regression for Descriptive Analysis 192

5-3-1 خصائص استخدام الانحدار في التحليل الوصفي 192

5-3-2 حالة دراسية رقم C5-7 باستخدام برنامج SPSS 192

5-4 استخدام تحليل المركبات الاساسية في التحليل الوصفي

Principal Components for Descriptive Analysis 200

5-4-1 خصائص استخدام المركبات الاساسية في التحليل الوصفي 200

5-4-2 حالة دراسية رقم C 5-8 باستخدام برنامج SPSS............ 200

الفصل السادس

أساليب بناء النماذج الاحصائية 211

Statistical Models Building Methods

6-1 تحليل الانحدار Regression analysis 211

6-2 تحليل الانحدار الخطي البسيط

Simple Linear Regression Analysis 211

6-2-1 معادلة الانحدار الخطي البسيط

Simple Linear Regression Equation 211

6-2-2 طريقة المربعات الصغرى وخواصها

Least Square Method.. 213

(1) خطوات طريقة المربعات الصغرى 213

(2) خواص مقدرات طريقة المربعات الصغرى 215

6-2-3 فرضيات نموذج الانحدار الخطي البسيط

Assumptions of Simple Linear Regression 218

6-2-4 اختبار فرضيات نموذج الانحدار الخطي البسيط 220

Test of SLRA assumptions ....

6 -2-5 حالة دراسية رقم C6-1 استخدام برنامج SPSS في الانحدار
الخطي البسيط 231
3-6 **تحليل الانحدار الخطي المتعدد Multiple Liner Regression** 239
6-3-1 معادلة الانحدار الخطي المتعدد وطريقة تقديرα, β's 240
6-3-2 معايير قياس كفاءة ومعنوية النموذج المتعدد 242
(1) معايير احصائية Statistical Crireria........................ 242
(2) معايير منطقية Logical Criteria........................... 243
(3) الفرضيات Assumptions 244
(4) **أختبار القوة التنبوئية للنموذج**
Predictive Power of the Model 250
(5) **الاختبار العملي للنموذج المطور**
Practical Testing of Developed Mode 251
6-3-3 طرق الانحدار الخطي المتعدد
Multiple Linear Regression Method 252
(1) طريقة شمول كافة المتغيرات All Possible Regression..... 252
(2) طريقة الاضافة المتتالية 252
Forward Selection Regression...........................
(3) طريقة الحذف التنازلي 252
Backward Elimination Selection Regression ....
(4) طريقة الخطوات المتتالية
Stepwise Selection Regression 253
6-3-4 حالة دراسية رقم C_{6-2}استخدام برنامج SPSS في تحليل الانحدار
المتعدد.. 253
4-6 **تحليل حساسية النموذج Sensitivity Analysis**.................. 263
6-4-1 مفهوم وخصائص تحليل الحساسية 263
6-4-2 حالة دراسية رقم C_{6-3} نماذج لاسلوب وضع
سينورياهات وفقا لتحليل حساسية المتغيرات........................ 266

6-5 تحليل الانحدار غير الخطي
Non-Linear Regression Analysis 271
6-5-1 تحليل الانحدار غير الخطي البسيط
Simple Non-Linear Regression Analysis 272
6-5-2 حالة دراسية C6-4 استخدام برنامج SPSS في تحليل
الانحدار غير الخطي البسيط.. 280
6-5-3 تحليل الانحدار غير الخطي المتعدد
Multiple Non-Linear Regression Analysis 286
(1) معادلة الانحدار التربيعية Quadratic Reg. Equation 286
(2) معادلة الانحدار التكعيبي Cubic Reg. Equation 289
6-5-4 حالة دراسية رقم C_{6-5} استخدام برنامج SPSS في تحليل
الانحدار غير الخطي المتعدد.. 290
6-6 تحليل المركبات الاساسية Principal Component Analysis .. 294
6-7 تحليل دالة التمييز Discriminent Analysis 295
6-7-1 خصائص واستخدامات دالة التمييز...... 295
6-7-2 حالة دراسية C_{6-6} استخدام برنامج SPSS في تحليل دالة التمييز.... 296

الفصل السابع

الدمج بين طريقتي تحليل الانـحدار وتحليل المركبات الاساسية في بناء النماذج

Model Building of Merging Regression & Principal Component Analyses

305

7-1 مقدمة Introduction 305
7-1-1 تحليل الانحدار Regression Analysis 306
7-1-2 تحليل المركبات الاساسية
Principal Component Analysis 307
7-1-3 الدمج بين تحليل الانحدار وتحليل المركبات،
and Principal Component Analyses.Emerging Reg 308

7-2 حالة دراسية C_{7-1} في استخدام طريقة الدمج بين الانحدار والمركبات 309

الفصل الثامن

اختبار الفروض Hypothesis Testing 317

8-1 مقدمة Introduction 317
8-1-1 مفهوم الفروض Hypothesis Definition 317
8-1-2 انواع اخطاء الفروض 319
8-1-3 اختبار من جانب ومن جانبين One & Two Sides tests 322
8-2 اختبار – ت T-test 324
8-2-1 الاختبار الاحادي One Sample test 324
(1) خصائص وأجراءات الاختبار الاحادي 324
(2) حالة دراسية رقم C_{8-1} في انجاز الاختبار الاحادي باستخدام برنامج SPSS 326
8-2-2 الاختبار لعينتين مستقلتين Samples Test Independent 330
(1) خصائص واجراءات اختبار الفروق بين مجتمعين مستقلين 330
(2) حالة دراسية رقم C8-2 في انجاز اختبارالفروق بين مجتمعين مستقلين باستخدام برنامج SPSS 337
8-2-3 اختبار المقارنات الزوجية Paired Samples Test 343
(1) خصائص واجراءات اختبار المقارنات الزوجية 343
(2) حالة دراسية رقم C8-3 في انجاز أختبار المقارنات الزوجية باستخدام برنامج SPSS 364
8-3 اختبار مربعات كاي $\chi2$ - test (Chi-Square) 348
8-3-1 خصائص اختبار التجانس واجراءات حسابه 348
8-3-2 حالة دراسية رقم c8-4 في انجاز أختبار $\chi2$ للتجانس باستخدام برنامج SPSS 349

8-4 **تحليل التباين Analysis of Variance** **353**
8-4-1 **خصائص تحليل التباين واجراءات حسابه.** **353**
8-4-2 تحليل التباين بمعيار واحد
One-Way Analysis of Variance **356**
(1) حالة تساوي حجوم العينات............................ **356**
(2) حالة عدم تساوي حجوم العينات............................ **359**
(3) حالة دراسية رقم C8-5 في انجاز تحليل التباين بمعيار واحد باستخدام برنامج SPSS **361**
8-4-3 تحليل التباين بمعيار واحد مع اكثر من مستوى واحد للمجموعة الواحدة Nested Analysis of Variance............................ **368**
8-4-4 تحليل التباين بمعيارين
Two Ways Analysis of Variance **373**
(1) خصائص واجراءات تحليل التباين بمعيارين **373**
(2) حالة دراسية رقم C8-6 لتحليل التباين بمعيارين باستخدام برنامج SPSS **377**
الملاحـــق **387**
ملحق 1.2 **387**
مقطع من التصنيف القياسي الدولي للانشطة الاقتصادية ISIC
ملحق 2.2 **389**
مقطع من التصنيف القياسي الدولي للتعليم ISCE
ملحق 3.2 **390**
مقطع من التصنيف القياسي الدولي للمهن ISCO
ملحق 1.3 **391**
نموذج لجدول الارقام العشوائية Random Numbers Table
ملحق2.3 **392**
المساحة تحت التوزيع الطبيعي القياسي الواقعة بين المتوسط وZ

ملحق 3.3 393
دالة التوزيع الطبيعي التجميعي لاحتمال المتغير العشوائي $N(0,1)$
ملحق 4.3 394
قيم z الجدولية الموزعة طبيعيا $N(0,1)$ عند مستويات معنوية مختلفة
ملحق 5.3 395
قيم t الجدولية عند مستويات معنوية ودرجات حرية $v = n - 1$
ملحق 6.3 396
قيم مربع كاي χ^2 عند مستويات معنوية α ودرجات حرية v
ملحق 7.3 397
قيم f الجدولية عند عدد من مستويات المعنوية ودرجات الحرية v_1 و v_2 عند مستوى معنوية 0.05
ملحق 1.6 398
قيم داربن- واتسون الجدولية Durban – Watson Table
المراجع 401
- الكتب العربية 401
- الكتب الانكليزية 402
- المجلات العلمية 404
المؤلف في سطور 405

قائمة الحالات الدراسية
List of Case Studies

- **حالة دراسية رقم** C_{1-1} : استخدام برنامج SPSS في تحويل المعطيـات الغـير رقمية الى قمية
- حالة دراسية رقم C_{1-2} : استخدام برنامج SPSS في الحصول علـى معطيـات على مستوى المفردة والمستوى التجميعي
- **حالة دراسية رقم** C_{2-1} : تصميم أستبانة تتكون من ثلاثة اجزاء
- **حالة دراسية رقم** C_{2-2} : تصميم أستبانة تتكون من جزئين
- **حالة دراسية رقم** C_{3-1} : في توزيع المعاينة Sampling Distribution (إثبات مدى تكافؤ إحصاءات العينة مع معالم المجتمع المسحوبة منه العينة)
- **حالة دراسية رقم** C_{3-2} : في تصميم عينة لنشاط الخدمات المالية
- **حالة دراسية رقم** C_{3-3} : في تحديد حجم عينة مسحوبة من مجتمع مجهول المعالم
- **حالة دراسية رقم** C_{4-1} : في استخدام الانحدار في التقدير والتعويض
- **حالة دراسية رقم** C_{5-1} : استخدام برنامج SPSS في المقاييس الوصفية
- **حالة دراسية رقم** C_{5-2} : استخدام برنامج SPSS في الارتباط البسيط
- **حالة دراسية رقم** C_{5-3} : استخدام برنامج SPSS في الارتباط المتعدد
- **حالة دراسية رقم** C_{5-4} : استخدام برنامج SPSS في الارتباط الجزئي
- **حالة دراسية رقم** C_{5-5} : استخدام برنامج SPSS في ارتباط الرتب
- **حالة دراسية رقم** C_{5-6} : استخدام برنامج SPSS في ارتباط التوافق
- **حالة دراسية رقم** C_{5-7} : توظيف الانحدار في التحليل الوصفي
- **حالة دراسية رقم** C_{5-8} : توظيف تحليل المركبات في التحليل الوصفي
- **حالة دراسية رقم** C_{6-1} : استخدام برنامج SPSS في الانحدار الخطي البسيط

- **حالة دراسية رقم** C_{6-2}: استخدام برنامج SPSS في تحليل الانحدار الخطي المتعدد
- **حالة دراسية رقم** C_{6-3} : نماذج لاسلوب وضع سينورياهات وفقا لتحليل حساسية المتغيرات
- **حالة دراسية رقم** C_{6-4} : استخدام برنامج SPSS في تحليل الانحدار غير الخطي البسيط
- **حالة دراسية رقم** C_{6-5} : استخدام برنامج SPSS في تحليل الانحدار غير الخطي المتعدد
- **حالة دراسية رقم** C_{6-6} : استخدام برنامج SPSS في تحليل دالة التمييز
- **حالة دراسية رقم** C_{7-1} : استخدام طريقة الدمج بين المركبات الاساسية والانحدار
- **حالة دراسية رقم** C_{8-1} : استخدام برنامج SPSS لانجاز الاختبار الاحادي
- **حالة دراسية رقم** C_{8-2} : استخدام برنامج SPSS لانجاز اختبارالفروق بين مجتمعين مستقلين
- **حالة دراسية رقم** C_{8-3} : استخدام برنامج SPSS في أختبار المقارنات لزوجية
- **حالة دراسية رقم** C_{8-4} : ستخدام برنامج SPSS لانجاز أختبار χ^2 للتجانس
- **حالة دراسية رقم** C_{8-5} : استخدام برنامج SPSS لانجاز تحليل التباين بمعيار
- **حالة دراسية رقم** C_{8-6} : استخدام برنامج SPSS لتحليل التباين بمعيارين

المقدمة Introduction

ان هذا الكتاب يجمع بين الخبرة الاكاديمية والعملية في مجالات البحوث والدراسات، روعي في اعداده توخي التبسيط والتيسير والتسلسل المنطقي نحو متطلبات اعداد الدراسات والبحوث. معززين كل ما يذكر في الكتاب بحالات دراسية بامل ان تكون وافية قدر الامكان، مبتعدين عن الخوض في مفاصل نظرية غير ضرورية وايلاء الاهتمام والتركيز على المواضيع التطبيقية والخطوات اللازمة لاستخدام برنامج SPSS والحصول على مخرجاته .

فكان الاستهلال في **الفصل الاول** باعطاء صورة عامة عما تهدف اليه الدراسات والبحوث مع تسليط الضوء على مفاهيم بعض المفاصل المرتبطة بالعمل البحثي. وتضمن **الفصل الثاني** تحديد احتياجات البحث الى المعطيات وكيفية وضعها في استبانة أحصائية مع تناول اسلوب تصميم الاستبانة وشروط صياغة اسئلتها في.

لناتي بعد تصميم الاستبيانة الى تحديد عدد الاستبانات اللازم ملئها المتمثل بتحديد حجم العينة ونوع العينة المطلوب تطبيقها واسلوب الاختيار لوحداتها وكيفية تعميم نتائج العينة على المجتمع الكلي ولكل نوع من العينات التي استخدامها وهو ماتم تناوله في **الفصل الثالث**.

تم العروج الى **الفصل الرابع** ليتم فيه تناول اسلوب تدقيق ما نقوم بجمعه من معطيات وكيفية التاكد من صحتها، والى اسلوب معالجة ماهو مفقود وغير مكتمل في هذه المعطيات لنكون جاهزين لمرحلة التحليل.

تلى ذلك الانغماس في مرحلة التحليل بدءا من الوصف والتفسير وادواته واسلوبه في **الفصل الخامس** ولغاية بناء النماذج الاحصائية المتقدمة،

واسلوب تحليل حساسيتها وكيفية استخدامها عمليا في التخطيط واتخاذ القرارات وهو ماتضمنه **الفصل السادس**.

وفي الفصل السابع تم عرض اسلوب جديد قديم في بناء النماذج، يتمثل بدمج اداتين لتصبح واحدة، وهو موضوع جديد في امكانية تطبيقه، قديم في طرحه لطول عملياته التحليله انذاك اي قبل التوسع باستخدام الحاسوب وقبل تيسيرات برنامج SPSS الا انه ظل محدود التطبيق تهيبا من الجديد وخوفا من المجهول.

لننتهي في **الفصل الاخير** بموضوع حيوي بحاجه لاستخدامه دائما وفي كافة المجالات وهو الاختبارات بما في ذلك تحليل التباين، ليكون موضوع الفصل الثامن .

املا ان ياتي ما اقدمنا عليه بالفائدة للباحثين وللعمل البحثي باعتباره الاسلوب المجدي لمن يريد حقا التغيير نحو الافضل في مجتمع اليوم، والله الموفق.

د. عبدالحميد عبدالمجيد البلداوي

beldawin@yahoo.ca

الفصل الاول

تمهيد ومفاهيم اساسية

Preface and Basic Definitions

1-1 اهمية البحوث العلمية Scientific Research Importance

أن البحوث هي ركيزة القرارات الصائبة التي تتخذها الحكومات او مؤسسات القطاع الخاص او الاشخاص سواء في مواجهة المشاكل والتغلب عليها او في سعيها للتطور او الاصلاح على مختلف ميادين الحياة، بحيث اصبحت في عصرنا الراهن سمة تقترن بقوة المجتمعات وتقدمها. فطبقا للكتب السنوية الصادرة عن منظمة اليونسكو، نجد ان دولا متطورة مثل اليابان والولايات المتحدة الامريكية تنفق ما يقارب 3٪ من اجمالي ناتجها المحلي على العمل البحثي، ويعمل بما معدله بحدود 5000 باحث لكل مليون نسمة في هذه الدول وذلك سعيا من اجل تحقيق :

1-1-1 الابتكار ومواكبة التطورات السريعة والشاسعة للعلوم والتكنولوجيا ومواجهة اثارها.

1-1-2 تهيئة التعليم لتكون مخرجاته قادرة على ادراك الحاجة الى الوعي بالتغير وتقدير الافكار المحتملة والاستعداد لها والتكيف على طرق الاداء والتقييم لما يطرأ من مستجدات.

1-1-3 ايجاد الحلول والبدائل الجديدة المبتكرة في معالجة ومواجهة المشاكل الاجتماعية والاقتصادية ووضع التصورات العلمية للاثار التي يمكن ان تحدثها القرارات على المستقبل

1-1-4 تطوير اساليب وطرق انتاج وتوزيع السلع والخدمات كميا ونوعيا، لاجل مواجهة المنافسة المتزايدة في غزو الاسواق الحاد.

1-2 مفهوم البحث العلمي وخصائصه

Scientific Research Definition and Properties

ان البحث العلمي هو وسيلة يتم بواسطته دراسة ظاهرة او مشكلة ما، للوقوف على العوامل التي ادت الى وقوعها، او للوصول الى علاجها او لتطويرها او لادخال تعديل عليها.

وقد عرف البعض البحث من انه "طريقة منظمة لاكتشاف حقائق جديدة بعد التثبت من حقائق قديمة والوقوف على القوانين التي تحكمها". كما عرف اخرون البحث من انه "محاولة لاكتشاف المعرفة والتنقيب عنها، وتطويرها وفحصها، ثم عرضها بذكاء وادراك لمواكبة التقدم والمساهمة ايجابيا في مسيرة التطور" (غرايبة واخرون ،2002).

وان اقتران البحث بالعلمية متاتي من الاستخدام المنظم لاساليب وادوات تحكمها معايير وفرضيات ذات اسس نظرية ومنطقية وتتميز بالخصائص التالية :

1-2-1 الموضوعية والابتعاد عن التاثير الشخصي ، Objectivity

اي توقع الحصول على نفس النتائج لو تم دراسة ظاهرة ما من قبل عدة جهات او عدة باحثين وباتباع نفس المنهجية ،

1-2-2 الاستناد الى الحقائق ، Facts

اي ان تكون الظاهرة قابلة للملاحظة مع امكانية التحقق من نتائج البحث في اي وقت، فمثلا لو قادنا الاستدلال الى ان الطلب يزداد عند تخفيض الاسعار، عندها لابد ان يكون الدليل واضحا عند المقارنة بين حجم

مبيعات سلعة ما عند بيعها باسعار منخفضة مع حجم مبيعات نفس السلعة عند بيعها باسعار مرتفعة،

1-2-3 التعميم، Generalization

اي القدرة على تعميم نتائج العينة على المجتمع الذي سحبت منه، وهذا يعني ان تكون العينة التي تم الاعتماد عليها في الدراسة قد تم سحبها عشوائيا وبذلك فهي ممثلة لخصائص المجتمع تمثيلا صحيحا ،

1-2-4 التوافق بين الاستنباط والاستقراء،

Induction & Deduction Motching

فالاستقراء يعني ملاحظة الظواهر والتوصل الى تعميمات حولها طبقا للمعطيات، اما الاستنباط فيبدأ بالنظريات التي تستنبط منها الفرضيات ثم البحث عن المعطيات لاختبار صحة هذه الفرضيات، وهو مشابه لما يطلق عليه في التحليل الاحصائي بالمعايير المنطقية ،فمثلا نلاحظ حصول زيادة الطلب على خدمة او سلعة ما عندما ترتفع جودتها، وعليه فان اختبار المعطيات يجب ان يكون متوافقا وهذا الاستقراء، اي ان تاتي اشارة متغير الجودة باشارة موجبة والا تم رفض النتيجة ،

1-2-5 القدرة على بناء التنبؤات Predictive Ability

من خلال العوامل التي لها تاثير على الظاهرة تحت الدراسة ضمن حدود درجة الدقة المستهدفة، لان الفشل في ذلك يعني عدم استيفاء الشروط والمعايير العلمية في النموذج المستخدم لبناء التنبؤات.

1-2-6 تنوع سبل انجازالبحث العلمي ،

(1) البحث النظري Theoritical Research، وهو الذي يتم اجراءه بهدف المعرفة وتطوير المفاهيم النظرية او توضيح غموض

يحيط بظاهرة ما من دون النظر الى تطبيق نتائج البحث كما هو الحال في بحوث الرياضيات البحتة Pure Mathemtics مثلا.

(2) البحث التطبيقي **Applied Research**، والذي يستهدف ايجاد حل لمشكلة قائمة او التوصل لحل تطويري، وهذا النوع من البحوث يعتمد على دراسات ميدانية او تجارب مختبرية، والتكيف على تطبيق نتائجه عمليا، كما هو الحال في بحوث الانتاج والتسويق وغيرها.

(3) **بحث مزيج بين النظرية والتطبيق، Mixture of Theory & Application Research**، فالبحث النظري قد يفتح افاقا جديدة للبحث العلمي وتقوم عليه بحوث تطبيقية في ميادين عدة كالهندسة والفلك وما شابه ،مما يساعد على توجيه وتطوير مسارات البحوث التطبيقية بصورة مباشرة او غير مباشرة.

1-3 المعطيات الاحصائية وانواعها

1-3-1 المعطيات الكمية Quantitative data

وهي التي تعبر بشكل رقمي عن ظاهرة معينة، ويطلق عليها احيانا بالمعطيات المقاسة measured data وتمثل اية نشاط او فعالية على وفق المقدار المنجز، فنقيس الانتاج بالطن او الكيلو او المتر واجزاءه وما شابه، والتعبير عن السعر بالدينار او الدولار او الدرهم واجزائها وعن الزمن بالساعة والدقيقة ...... الخ. ان هذا النوع من المعطيات يعبر عن ظروف وخصائص اية سلعة او خدمة او ظاهرة كما هي عليه من دون اجتهاد او وجهة نظر.

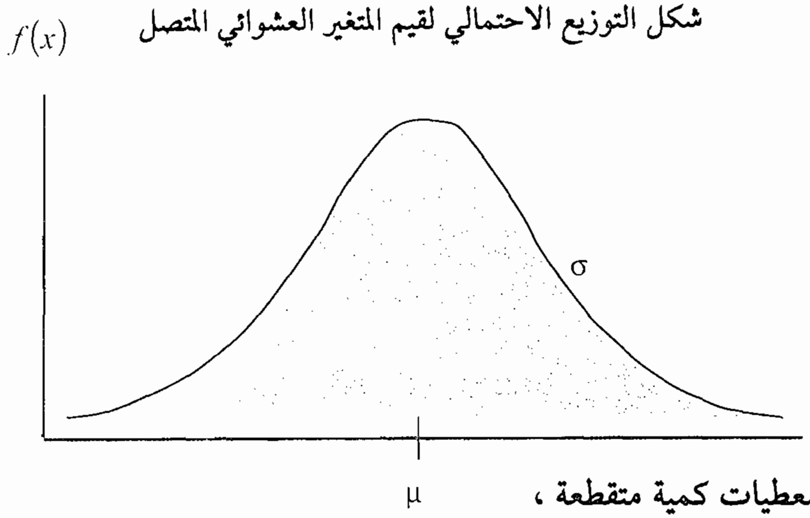
(1) معطيات كمية متصلة ،

عندما تشتمل قيم هذه المعطيات على كسور كما في حالة الاسعار او الاطوال او الاوزان مثلا يطلق عليها بالمتغيرات المستمرة او المتصلة

continuous variables، فان احتمال هذه القيم تشكل مساحة تحت المنحني وكما مبين في الشكل رقم (1-1)، وهذا النوع من المعطيات يسمح باستخدام الاساليب الكمية للتحليل التي تشترط استيفاء فرضية التوزيع الطبيعي واختبار جودة نتائجها.

شكل بياني رقم (1-1)

شكل التوزيع الاحتمالي لقيم المتغير العشوائي المتصل

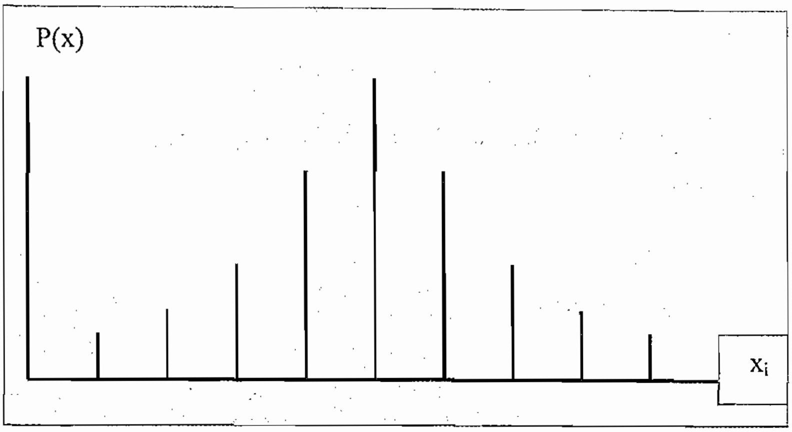


(2) معطيات كمية متقطعة ،

اما عندما تكون قيم المعطيات الرقمية عبارة عن اعداد صحيحة من دون كسور كما في حالة عدد الطلبة او عدد العاملين مثلا فتسمى بالمتغيرات المتقطعة discrete variables حيث يكون تمثيلها بيانيا عبارة عن نقاط منفصلة مما يتعذر تشكيل مساحة متصلة بين قيمها لصعوبة قياس عرض كل من هذه الاعمدة وبذلك تكون مساحتها مساوية للصفر وكما مبين في الشكل رقم (1-2)، مما يستوجب التخلص من مديات الفئات باستخراج ما يسمى بالحدود الحقيقية للفئات او بنهايات الفئات (النهاية الدنيا والنهاية العظمى) لاجل التواصل بين الفئات وبالتالي التمكن من ايجاد منحنى طبيعي تقريبي للقيم المتقطعة .

شكل بياني رقم (2-1)

الشكل العام للتوزيع الاحتمالي لقيم المتغير العشوائي المتقطع



2-3-1 المعطيات النوعية Qualitative data

وهي المعطيات التي تصف الظاهرة المعنية بشكل غير رقمي كالنوع (ذكور- اناث) والتحصيل الدراسي (دكتوراه- ماجسير- بكلوريوس ...الخ)، كما ويمكن تنظيم وحدات الظاهرة حسب اشتراكها في الصفة مثل ممتاز، جيد جدا، جيد، ... الخ. او قد تكون قيمة البيان تمثل راي الشخص المبحوث وقناعته، وهذه المعطيات تساعد في حل العديد من المشاكل الاجتماعية والاقتصادية كون الاشخاص الذين يدلون بارائهم يعتمدون عليها في اتخاذ قراراتهم عمليا. وهي على نوعين رئيسين هما:

(1) المتغيرات الاسمية Nominal Variables وهي المتغيرات التي لايمكن ترتيبها تصاعديا او تنازليا، لذلك يكون ترميزها Coding من دون معنى كمي لان ترتيب مواقع اصناف او فئات المتغير ياتي من دون افضلية فعند اعطاء الرمز 1 للذكور و 2 للاناث مثلا لايعني ان الرمز 2 يساوي

ضعف الرمز 1 للذكور، لانه بالامكان ترتيب الاناث قبل الذكور ايضا وبالتالي يكون الرمز 1 للاناث والرمز 2 للذكور .

(2) المتغيرات الترتيبية Ordinal Variables وهو المتغيرات التي يمكن ترتيب مستوياتها او فئاتها ترتيبا تصاعديا او تنازليا، لكن لايمكن تحديد مقدار الفروق او المسافات بدقة بين هذه المستويات او الفئات، فعندما يتكون المتغير من ثلاث مستويات مثلا هي عالي - متوسط - ضعيف، فالاجابات المحتملة ستصف الحجم النسبي وتمكننا فقط من معرفة ان عالي هي اكبر من متوسط ولكن لانستطيع معرقة مقدار حجم الفرق بين عالي و متوسط او بين متوسط وضعيف وهكذا .

وبذلك فان هذه المعطيات تكون بحاجة الى تحويلها الى قيم كمية للتمكن من اخضاعها للتحليل، وتتم عملية التحويل من خلال اعتماد نظام الدرجات scaling system، وهناك نمطين من آليات نظام الدرجات هما:

- **النمط ذات البعد الاحادي uni-dimensional scaling**، الذي بموجبه يفضل ان يكون تقسيم مستوى اهمية المتغير الى عدد فردي كأن يكون 3 مستويات او 5 او 7 الخ وحسب درجة الدقة المستهدفة وطبيعة المتغيرات، لتصبح نقطة الوسط هي 2 في حالة 3 مستويات و 3 في حالة 5 مستويات وهكذا. فمثلا في حالة تحديد مستوى جودة سلعة ما بـ 5 مستويات هي ردئ وتعطى له القيمة 1 ومقبول وتعطى له القيمة 2 و 3 لمستوى جيد و 4 لجيد جدا والقيمة 5 لمستوى ممتاز .

- **النمط الثاني من الاليات هو ذو الابعاد المتعددة multidimensional scaling**، والذي فيه يستمر السؤال بعد الاجابة الاولى فياتي سؤال ثاني يتعلق بالاجابة الاولى، فاذا افترضنا بان الاجابة

جاءت من ان السلعة رديئة فياتي السؤال اللاحق عن سبب كون السلعة رديئة، او الطلب من المبحوث تقديم مقترح او ابداء ما يراه مناسبا لتحسين السلعة لكي تكون ممتازة من وجهة نظره، وقد يتبع ذلك اسئلة اخرى تتعلق بذات الموضوع وهكذا.

3-1 النماذج الاحصائية مفهومها وأنواعها

Statistical Models Definition and Classification

النموذج الاحصائي (او الرياضي) هو عبارة عن استخدام الاساليب الاحصائية والرياضية لمعالجة عدة مراحل تحليلية، في ضوء مجموعة محددات احصائية ومنطقية وفرضيات قياسية، من اجل بناء اداة علمية يطلق عليها نموذج يضم المتغيرات الاساسية ذات الصلة بالظاهرة تحت الدراسة. بكلمة اخرى فان النموذج هو معادلة رياضية تضم العوامل المرتبطة بظاهرة ما وترينا درجة تاثير كل من هذه العوامل ومعنويتها. وتاتي اهمية بناء النماذج لمساعدة متخذ القرار او المخطط على فهم ما يمكن ان يحصل للظاهرة قبل الاقدام على عملية التغيير او التطوير وقبل صرف الاموال وبذل الجهود. ويعتمد تصنيف النماذج على معيار التصنيف المستخدم في بناؤها وكما يلي:

1- 4- 1 تصنيف النماذج حسب نوع المعطيات المستخدمة

فقد يتعلق التصنيف بنوع المعطيات التي يتم توظيفها في بناء النموذج، فان كانت المعطيات الموظفة وضعية (Situational Data) سميت بالنماذج الوضعية، وان تم توظيف معطيات استطلاع الراي او الاعتقاد (Attitudinal Data) اطلق عليها نماذج الراي، او بالنماذج السلوكية (Behavioral) اذا ما كانت المعطيات سلوكية، وهكذا .

1- 4- 2 تصنيف النماذج حسب صيغة المشاهدات

اي ان ياتي تصنيف النماذج وفقا لطبيعة المشاهدات (Observation)، فتدعي بالنماذج التجميعية (Aggregate Models) عندما تكون المشاهدة عبارة عن معدل لعدد من المشاهدات (وحدات العد)، او بنماذج المفردة (Disaggregate Models) اذا كانت عملية التوظيف هي القيمة المفردة المباشرة لوحدة المشاهدة .

1- 4- 3 تصنيف النماذج وفقا للهدف من البحث

وقد يكون التصنيف وفقا للهدف من استخدام النموذج، عندها تدعى نماذج تنبوئية (Predictive Models) واخرى تفسيرية او وصفية (Explanation or Descriptive Models) او نماذج سيطرة (Controlling Models) وهكذا.

1- 4- 4 تصنيف النماذج حسب نوع المعادلة المستخدمة

او ياتي التصنيف حسب المعادلة المستخدمة في عملية بناء النموذج، فتصنف الى نماذج انحدار (Regression Models)، ونماذج أحتمالية (Probabilistic Models) أو نماذج محددة (Deterministic Models)...الخ.

1- 5 الاهداف العامة للبحوث

1- 5- 1 الوصف والتفسير Description and Explanation

والوصف يعتبر الخطوة الاولى في تحقيق الهدف النهائي للبحث، وهو يعني شرح المعطيات التي يتم جمعها وتبويبها والوقوف على خصائصها مع سرد مبررات اعتمادها. اما التفسير فيعني الكشف والاستدلال على

الاسباب التي ادت الى حدوث الظواهر بالاعتماد على المقارنة والربط بين العناصر المختلفة للتوصل الى معرفة هذه الاسباب. كمحاولة للكشف عن الاسباب المؤدية الى ارتفاع معدل الجرائم او اسباب وقوع حوادث الطرق مثلا.

ويشمل ذلك الظواهر الاقتصادية والاجتماعية، والمسائل المتعلقة بالتكنولوجيا والطب والهندسة والتجارب المختبرية في علوم الفضاء والفلك والجولوجيا والزراعة وعلم الحيوان وغيرها العديد التي تحتاج لدقة عالية، ويكون قياس المتغيرات فيها باهض التكاليف مما تتطلب اللجوء الى استخدام التحليل الوصفي حصرا توخيا لواحد او اكثر من الاهداف التالية:

(1) معرفة مكنونات العوامل المؤثرة على ظاهرة ما، وطبيعة اتجاهاتها وفحص مسببات هذا التاثيرودرجته واتجاهه.

(2) المقارنة بين اهمية المتغير بالنسبة لظاهرة ما وحدود العناية المالية المستحقة لقياسه في استراتيجيات وخطط المؤسسة او المنظمة، ليتسنى البحث عن البدائل الممكنة في حالة كانت كلفته باهظة او صعوبة في الحصول عليه.

(3) لكشف المتغير وتفسيره علاقة مباشرة في التوصول لحلول مشكلات قائمة يجري البحث في ايجاد حلول لمعالجتها.

وبذلك يتركز التحليل هنا على وصف ظاهرة ما والكشف عن خصائصها او اتجاه تطورها سلبا او ايجابا، ويتم ذلك من خلال عدة طرق منها تحليل خصائص المعطيات الاحصائية المتمثلة بالنزعة المركزية بمختلف مقاييسها (المتوسطات) او المقاييس غير المركزية كالعشير والربيع والمئين،

وكذلك باستخدام مقاييس التشتت التي يقصد بها حالة انتشار المعطيات حول المتوسط لمعرفة حالة التشابه والاختلاف .

كما ويتم ايضا الاستعانة بتحليل الارتباط ليتسنى معرفة ان كانت هناك علاقة بين متغيرين او مجموعة متغيرات من عدمها ودرجة هذه العلاقة واتجاهها ففي حالة الارتباط البسيط يقال ان العلاقة موجبة اذا كانت قيم الظاهرة (المتغير التابع dependent variable) تميل نحو الارتفاع كلما ازدادت قيم المتغير المستقل independent variable. اما اذا كانت قيم الظاهرة تميل نحو الانخفاض كلما ازدادت قيم المتغير المستقل عندها يقال ان العلاقة سالبة.

وبالاضافة للادوات التحليلية اعلاه بالامكان ايضا توظيف تحليل الانحدار لغرض تفسير او وصف ظاهرة ما، و يتم ذلك من خلال شمول المعادلة على جميع او اغلب المتغيرات المرشحة التي يعتقد بان لها علاقة مع الظاهرة بغض النظر عن درجة معنويتها، وعادة لايمكن في مثل هذه الحالة استخدام هكذا معادلة لاغراض بناء توقعات او لاهداف اخرى غير الوصف و التفسير لانها في الغالب تضم بعض او العديد من المتغيرات التي لا تضيف للمعادلة نسبة ملموسة او معنوية في تفسير التباين نتيجة ضعف تاثير بعض هذه المتغيرات على الظاهرة وفقا للمعاييرالاحصائية او بسبب الترابط شبه التام فيما بين البعض من المتغيرات المستقلة التي تضمها المعادلة مما يكون له تاثير سلبي على استيفاء نموذج الانحدار للفرضيات.

كما وان تحليل المركبات الاساسية principal component analysis الذي هو من الادوات الاحصائية المتقدمة، يعتبر احد الاساليب الوصفية ايضا وهو طريقة تستخدم مع المعطيات ذات المتغيرات المتعددة، حيث يقوم

بتجميع كل مجموعة من المتغيرات المترابطة خطيا وتحميلها loading في احد المركبات الاساسية ليرينا درجة وقوة علاقة المركبة المعنية مع مجموعة المتغيرات وفقا لمعيار التباين الذاتي eignvalues، وذلك بالاعتماد على مصفوفة الارتباط للمتغيرات التي تعتبر المرحلة الاساس في عملية تحليل المركبات.

1- 5- 2 بناء تقديرات و توقعات

ويتطلب هذا الهدف توفر درجة دقة و معنوية عالية في النماذج التي يتم تطويرها، حيث تخضع لمعايير وفرضيات متعددة قبل قبولها واستخدامها في بناء تقديرات و توقعات مستقبلية، ولا تتضمن هذه النماذج الا المتغيرات ذات التاثير المعنوي على الظاهرة ويتم ذلك من خلال اما توظيف احدى طرق الانحدار المتعلقة باختيار افضل طاقم متغيرات مستقلة او باستخدام تحليل المركبات الاساسية.

1- 5- 3 للسيطرة والتحكم

ويعني التحكم في العوامل المؤثرة على الظاهرة سواء في حالة وقوعها او في منعها من الوقوع، ويتم ذلك من خلال بناء النماذج وتطويرها و القيام بتحليل لكل من العوامل التي يتم كشفها على ان لها تاثير على الظاهرة لغرض الاستدلال على الحال الذي سيؤول اليه المتغير التابع (الظاهرة) من خلال افتراضات متعددة تتعلق بكل من هذه المتغيرات المستقلة التي يتضمنها النموذج وباستخدام تحليل مرونة كل من هذه العوامل (المتغيرات المستقلة) والحدود التي يمكن ان يذهب اليها في عملية التطوير وهو ما يطلق عليه بتحليل الحساسية Sensitivity analysis. ان هذا النوع من التحليل يتيح لمتخذ القرار او المخطط اختبار عدة سيناريوهات ويوفر بدائل لسياسات مختلفة

ليختار من بينها ما هو افضل. بكلمة اخرى ان عملية تحليل الحساسية Sensitivity Analysis ترينا درجة استجابة الظاهرة للتغير الذي يمكن ان يطرأ على أي من المتغيرات المستقلة التي يتضمنها النموذج.ويمكن استخدام تحليل المرونات Elecitecity لقياس درجة مرونة المتغيرات المستقلة X_i وحدود قدرتها في المساهمة على تغييراو تطوير المتغير التابع Y .

ان لهذه المرحلة التحليلية اهمية خاصة حيث انها تمكن متخذ القرار او المخطط من استخدامها في السيطرة والتحكم في عملية التطوير المستهدفه وفقا للامكانيات المتاحة للمنظمة، وكما ذكرنا فمن خلالها نستدل على :

(1) مدى قدرة وحدود كل من المتغيرات التي يتضمنها النموذج في تحقيق عملية التطوير

(2) مدى تناسب الامكانيات المادية والبشرية (حجم الاستثمار) التي يتم توظيفها مع حجم التطوير او المردود المتوقع

وتتشابه الاجراءات المطلوبة هنا مع تلك التي تم سردها مع متطلبات نماذج بناء التوقعات المستقبلية وحاجتها الى مستوى عالي من الدقة، مع التركيز على ضرورة ومحاولة تضمين النموذج للمتغيرات التي تتميز بمرونة وتسمح في التحكم بها من قبل متخذ القرار كمتغيرالاجور او الاسعار او جودة الخدمات او البضاعة او سرعة واسطة النقل او نوعية المواد الاولية .... الخ بالاضافة الى امتلاكها التاثير المباشر على الظاهرة المتمثلة بالمتغير التابع، ويطلق على هذا النوع من المتغيرات policy variables. وبذلك فان التقديرات الدقيقة والكفوءة لمعاملات النموذج بالاضافة الى اختيار افضل طاقم متغيرات يعكس العلاقات المهمة والمعنوية التي تكون هي الهدف الابرز عند بناء هذه النماذج.

1- 5- 4 اختبار الفروض

وهي واسعة الاستخدام و تهدف الى التحقق من مدى التجانس اوالاختلاف بين وحدات ظاهرة ما او بين ظاهرتين او اكثر بسبب فروق مكانية او زمنية او نوعية باستخدام ادوات التحليل الاستدلالي كتحليل التباين analysis of varaince، او اختبار – ت t-test، او مربعات كاي χ^2، والتي سيلي التطرق اليها في الفصل الثامن.

1- 6 منهجيات البحث العلمي

Scietific Research Methodologies

1- 6- 1 منهجية المفردة Disagreggate Methodology

وهي المنهجية التي تكون مشاهداتها (وحدات العينة) عبارة عن مفردة كالشخص او الاسرة او الشجرة او الحيوان مثلا، وتاخذ تسمية نماذجها من طبيعة المعطيات التي تتضمنها هذه النماذج والتي يمكن تصنيفها الى الانواع التالية :

(1) المعطيات السلوكية Behavioural Data

وهو ان الاساس النظري لاختيار السلعة او الخدمة يعتمد على السلوك الاقتصادي للفرد ولخصائصه، وهذا النوع من القرارات هو احتمالي بطبيعته، وانها تعزى الى نظرية الترشيد او العقلانية لسلوك الاختيار (theory of rational choice behaviour). فاذا افترضنا بان خصائص الفرد الاقتصادية والاجتماعية هي S وان خصائص السلعة او الخدمة هي X فان قرار الاختيار سيتوقف على دالة المنفعة Utility المتحققة من المقارنة بين اختياره وبين السلع والخدمات الاخرى المتاحة، اي ان تقييمه او تفضيله للسلعة او الخدمة i سيتوقف على خصائصه وعلى خصائص السلع او الخدمات الاخرى فيشعر بان :

$$U(X_j, S) > U(X_i, S)$$
$$\text{For } i \neq j \,;\, j = 1, 2, \ldots, j$$

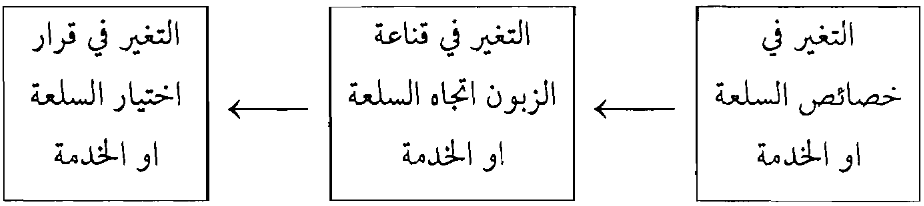
ومن الطرق الاحصائية التحليلية المناسب استخدامها مع هذا النوع من المعطيات في بناء هذه النماذج هي الدالة المميزةdiscriminante analysis اوprobit analysis او logit analysis والشكل العام لعلاقة هذه النماذج هو :

$$P = e^{u} / 1 + e^{u}$$

ويتم ايجاد قيمة U بدلالة المتغيرات المستقلة (Independent variables) يمكن استخدام طريقة الانحدار بتوظيف طريقة المربعات الصغرى least squear method او طريقة الدالة المفضلة maximum likeliood.

(2) معطيات الراي او الاعتقاد Attitudinal Data

وهي المعطيات التي تعتمد على اراء الاشخاص واعتقادهم اتجاه كل من الخيارات المتاحة امامهم من السلع والخدمات، وتاتي عملية الاعتقاد كحصيلة خبرة او ادراك ذاتي وعلى عوامل اقتصادية واجتماعية ترتبط بالشخص المبحوث respondent ويمكن اعتبار هذا النوع من المعطيات والنماذج هي الاكثر ملائمة في التخطيط ولاستطلاعات الاراء لاغراض التقييم للاداء وللتطوير، وتعتبر العلوم الاجتماعية والنفسية هي من اكثر الحقول استخداما ومساهمة في تطوير النماذج التي تستخدم هذا النوع من المعطيات. والمخطط التالي يوضح العلاقة بين موقف الزبون من التغير الذي يطرأ على خصائص السلعة او الخدمة ودرجة الرضا او القناعة.



وتمتاز نماذج هذه المعطيات بقابليتها العالية للاستجابة للمنتج او المخطط او متخذ القرار لكونها تقوم على فرضية ان ما يدلي به الزبون هو مطابق لما يفضل فعله مما يعطي الانطباع عن اعتماديتها وحقيقة تعبيرها لسلوك الزبون. وبذلك فان هذه النماذج تتلافى عيوب النماذج التجميعية.

ولتقريب مفهوم منهجية المفردة Disagreggate Methodology نسوق المثال (1-1) في ادناه :

مثال (1-1) : كانت نتائج استقصاء (مسح) شمل عينة عدد وحداتها $n = 31$ من ساكني مدينة عمان الاردنية، تناولت بعض خصائصهم وهي: منطقة السكن، العمر، الجنس، معدل الدخل الشهرى للاسرة، و ارائهم عن مستوى خدمات النقل العام (الباصات) في المدينة، والمتعلقة بـ: مستوى توفر النقل، مستوى نوعية وسائط النقل من ناحية الراحة والملائمة، مستوى اجور النقل، مدى ملائمة مواقع توفر خدمات النقل من حيث قربها لمواقع العمل والسكن، ودرجة الرضا العام عن الخدمة. وجاءت نتائج الاجابة على اسئلة الاستبانات كما مبين في جدول تفريغ المعطيات رقم (1-1) . والمطلوب تبيان كيفية التعامل مع المعطيات وفقا لمنهجية المفردة ومن ثم المنهجية التجميعية .

جدول (1-1)

جدول تفريغ الاجابات على اسئلة الاستبانة حسب تسلسل الاستبانة

منطقة السكن	العمر سنة	الجنس	الدخل دينار	توفر النقل	راحة النقل	اجور النقل	موقع النقلل	الرضا العام
جبيهة	45	ذكر	260	جيد	ممتاز	مقبول	مقبول	جيد
وادي السير	25	انثى	180	جيد جدا	جيد جدا	مقبول	سيئ	جيد

منطقة السكن	العمر سنة	الجنس	الدخل دينار	توفر النقل	راحة النقل	اجور النقل	موقع النقلل	الرضا العام
ابو نصير	51	انثى	260	ممتاز	جيد	سيئ	جيد	جيد
ابو نصير	19	انثى	500	جيد	ممتاز	جيد	جيد جدا	ممتاز
جبل الحسين	46	ذكر	310	جيد جدا	جيد جدا	مقبول	جيد	مقبول
صويلح	52	انثى	420	جيد	جيد	مقبول	مقبول	سئ
صويفية	22	ذكر	450	ممتاز	جيد	جيد	جيد جدا	جيد جدا
خلدا	34	ذكر	320	جيد جدا	جيد	جيد	سيئ	جيد
وادي السير	20	ذكر	210	جيد	جيد حدا	جيد جدا	جيد	جيد جدا
وادي السير	34	ذكر	260	حيد جدا	جيد	جيد	ممتاز	جيد
صويلح	32	انثى	400	جيد جدا	جيد جدا	مقبول	جيد	مقبول
صويلح	58	ذكر	310	ممتاز	جيد جدا	جيد	جيد جدا	جيد جدا
صويلح	21	انثى	200	جيد	ممتاز	جيد جدا	جيد	جيد
ابو نصير	38	ذكر	400	ممتاز	جيد	سيئ	جيد	سئ
خلدا	58	انثى	270	جيد جدا	جيد جدا	مقبول	جيد جدا	مقبول
جبل الحسين	43	ذكر	310	ممتاز	جيد جدا	مقبول	مقبول	مقبول

منطقة السكن	العمر سنة	الجنس	الدخل دينار	توفر النقل	راحة النقل	اجور النقل	موقع النقلل	الرضا العام
جبل الحسين	22	انثى	230	حيد جدا	مقبول	مقبول	جيد جدا	جيد
جبل الحسين	60	ذكر	280	ممتاز	جيد جدا	سيئ	جيد جدا	سئ
ابو نصير	34	ذكر	260	جيد جدا	جيد جدا	مقبول	جيد	مقبول
جبيهة	40	انثى	220	جيد	جيد	سيئ	جيد جدا	مقبول
صويلح	19	ذكر	330	ممتاز	جيد جدا	مقبول	جيد جدا	جيد
خلدا	51	انثى	550	مقبول	حيد	مقبول	جيد	جيد
جبيهة	27	ذكر	210	جيد	جيد	سئ	جيد	مقبول
جبل الحسين	38	ذكر	320	جيد جدا	جيد	مقبول	جيد	جيد
صويلح	31	ذكر	190	جيد	جيد	سئ	جيد جدا	حيد
ابو نصير	45	انثى	320	سئ	جيد جدا	مقبول	جيد	مقبول
وادي السير	50	ذكر	610	مقبول	مقبول	مقبول	جيد	مقبول
وادي السير	31	ذكر	200	جيد	جيد	مقبول	مقبول	مقبول
صويفية	48	انثى	430	جيد	جيد	جيد	مقبول	جيد
صويلح	21	انثى	650	جيد	جيد جدا	جيد	جيد	جيد
ابو نصير	27	ذكر	230	جيد	جيد جدا	مقبول	جيد	مقبول

ان صيغة جدول المدخلات لهذه المعطيات رقم (1-1) اعلاه يمثل المنهجية التي تعتمد المفردة، وبذلك فان عدد المشاهدات هو مساويا لعدد الاستبانات وهي 31. ولكي تصبح المعطيات النوعية التي تضمنها الجدول قابلة لعملية التحليل، يتطلب تحويلها الى معطيات كمية، ويتم ذلك كالاتي :

- ترميزكل من متغيرات: مستوى توفر النقل(Availability)، مستوى نوعية وسائط النقل من ناحية الراحة والملائمة (Comfort & Conveneince)، مستوى اجور النقل(Travel Cost)، مدى ملائمة مواقع توفر خدمات النقل من حيث قربها لمواقع العمل والسكن (Accesbility)، فقد قسمت الى 5 درجات، ويتم ترميز الاجابة على هذه المتغيرات بالقيم التالية: 1 لاجابة سيء، 2 لاجابة مقبول، 3 لاجابة جيد، 4 لاجابة جيد جدا، و 5 لاجابة ممتاز. اما بخصوص المتغير الاسمي وهو الجنس فاعطيت القيمة 1 للذكور و2 للاناث، وقيم هذه الرموز كما ذكرنا لاتدل على درجة الاهمية، حيث يمكن مثلا اعطاء 2 للذكوروالرمز 1 للاناث من دون ان يؤثر ذلك في درجة تفاعله التحليلية .

- ترميز منطقة السكن التي يتم توزيع المعطيات بموجبها، لتاخذ القيم التالية

جبيهة 1، وادي السير2، ابو نصير3، جبل الحسين 4، صويلح 5، صويفية 6 ، خلدا 7.

- فنحصل على الجدول رقم (1-2) التالي :

جدول رقم (2-1)

مدخلات معطيات المثال (1-1) بموجب منهجية التحليل التي تعتمد المفردة

رقم* المشاهدة	منطقة السكن	X01 العمر	X02 الجنس	X03 الدخل (دينار)	X04 توفر النقل	X05 راحة النقل	X06 اجور النقل	X07 موقع النقل	الرضا العام
01	1	45	1	260	3	5	2	2	3
02	2	25	2	180	4	4	2	1	3
03	3	51	2	260	5	3	1	3	3
04	3	19	2	500	3	5	3	4	5
05	4	46	1	310	4	4	2	3	2
06	5	52	2	420	4	4	2	2	1
07	6	22	1	450	5	3	3	4	3
08	7	34	1	320	4	3	2	1	3
09	2	20	1	210	3	4	4	3	4
10	2	34	1	260	4	3	3	5	3
11	5	32	2	400	4	4	2	3	2
12	5	58	1	310	5	4	3	4	4
13	5	21	2	200	3	5	3	3	3
14	3	38	1	400	5	3	1	3	1
15	7	58	2	270	4	4	2	4	2
16	4	43	1	310	5	4	2	2	2
17	4	22	2	230	4	3	2	4	3
18	4	60	1	280	5	4	1	4	1
19	3	34	1	260	4	4	2	3	2
20	1	40	2	220	3	3	1	4	2
21	6	19	1	330	5	4	2	4	3
22	7	51	2	550	1	3	2	3	3
23	1	27	1	210	2	3	1	3	2
24	4	38	1	320	4	3	2	3	3
25	5	31	1	190	3	3	1	3	3
26	3	45	2	320	1	4	2	4	2
27	2	50	1	610	2	2	2	2	2
28	2	31	1	200	3	3	2	3	2
29	6	48	2	430	3	3	3	3	3
30	5	21	2	650	3	4	3	4	3
31	3	27	1	230	3	4	1	4	2

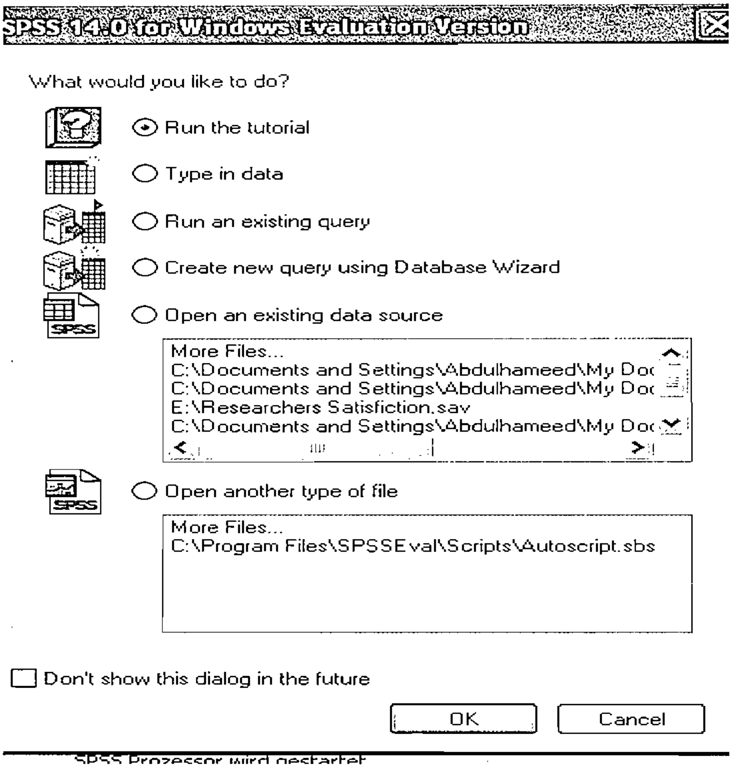
(2)حالة دراسية رقم C_{1-1}

في تحويل المعطيات الغير رقمية الى رقمية باستخدام برنامج SPSS

- الدخول الى البرنامج : start ⟹ program ⟹ SPSS
- تظهر لنا لوحة تحمل قائمة بالخيارات وكما مبين في الشكل البياني رقم (1.1) ليتم تاشير الملف المطلوب استخدامه او ان يكون الخيار هو لانشاء ملف جديد او باللجوء الى الامر الرئيسي File ومن ثم اختيار الامر الفرعي New في حالة انشاء ملف جديد او اختيار احد الملفات الموجودة مسبقا للعمل عليه .

الشكل بياني رقم (1.1)

لوحة قائمة الخيارات المتاحة لبدأ العمل مع برنامج SPSS



▪ تبدأ عملية انشاء ملف وادخال المعطيات في حالة استخدام اللوحة التي تحمل قائمة الخيارات من خلال التاشير على موقع Type in data المبين في الشكل رقم (1.1)، ومن ثم الكبس غلى ايقونة Ok الموجودة في اسفل القائمة فتظهر صفحة الجدول التي يتم فيها تدوين اسماء المتغيرات المزمع تبويب معطياتها المبينة في الشكل رقم (2.1) والتي تحمل عنوان Variable view المدونة في اسفل الجدول. كما يتم فيها ادراج المعلومات القاموسية المطلوبة بخصوص كل متغير معني وهي : نوع الترميز ويشار اليها بـ Type للاشارة ان كان المتغير رقمي numeric او اسمي String (وهنا لابد من التاكيد على ضرورة التاشير string ليتسنى قبول البرنامج تدوين اسماء المغيرات النوعية)، وعدد الخانات المطلوبة Width، عدد المراتب العشرية Decimals، وتعريف المتغير او عنوانه Label، القيمة Value، تعريف القيم المفقودة Missing values، عدد مراتب العمود Columns، تنسيق العمود Align، ونوع القياس Measure .

شكل بياني رقم (2.1)

يبين صفحة Variable View لتدوين المعلومات المتعلقة بالمتغيرات

1 : District jubaih

	District	Age	Gender	MFIncome	TAvailaS	TComConS	TCostS	TAccesibS	YGS
1	jubaih	45.00	1	260.00	g	d	a	a	g
2	wadisa	25.00	2	180.00	v	v	a	b	g
3	abunas	51.00	2	260.00	d	g	b	g	g
4	abunas	19 00	2	500.00	g	d	g	v	d
5	jabilh	46.00	1	310.00	v	v	a	g	a
6	jubaih	45.00	1	260.00	g	d	a	a	g
7	wadisa	25.00	2	180 00	v	v	a	b	g
8	abunas	51.00	2	260.00	d	g	b	g	g
9	abunas	19.00	2	500.00	g	d	g	v	d
10	jabilh	46.00	1	310.00	v	v	a	g	a
11	swalih	52.00	2	420.00	g	g	a	a	b
12	swifia	22.00	1	450.00	d	g	g	v	v
13	khalda	34.00	1	320.00	v	g	g	b	g
14	wadisa	20.00	1	210.00	g	v	v	g	v
15	wadisa	34.00	1	260.00	v	g	g	d	g
16	swalih	32.00	2	400.00	v	v	a	g	a
17	swalih	58.00	1	310.00	d	v	g	v	v
18	swalih	21.00	2	200.00	g	d	v	g	g
19	abunas	28.00	1	400.00	d	g	b	g	b
20	khalda	58.00	2	270.00	v	v	a	v	a
21	jabilh	43.00	1	310.00	d	v	a	a	a
22	jabilh	22.00	2	230.00	v	a	b	v	g

Data View / Variable View

▪ وعقب الانتهاء من تدوين اسماء المتغيرات والمعلومات القاموسية المتعلقة بها، يتم الكبس على ايقونة Data View المبينة في اسفل ذات الصفحة ايضا، ليظهر الجدول المبين في الشكل البياني رقم (3.1) الذي يتم فيه ادخال المعطيات ويجري ذلك بشكل متسلسل فكل صف (سطر) تعود معطياته لمشاهدة معينة (كان تكون استبانة اوسنة او وحدة زمنية او مكانية او شخص) وكل موقع (خانة) في السطر تعود لمتغير محدد، وفي حالة مصادفة معطيات مفقودة يترك مكانها خاليا ليتم معالجتها لاحقا بعد الانتهاء من عملية الادخال اما بتقديرها او تعويضها باحد اساليب التقدير او التعويض التي سيلي ذكرها.

شكل بياني رقم (3.1)

يبين صفحة Data view التي يتم فيها تدوين المعطيات عند انشاء الملف

32 : YGS

	District	Age	Gender	MFIncome	TAvailaS	TComConS	TCostS	TAccesibS	YGS
6	swalih	52.00	2	420.00	g	g	a	a	b
7	swifia	22.00	1	450.00	d	g	g	v	v
8	khalda	34.00	1	320.00	v	g	g	b	g
9	wadisa	20.00	1	210.00	g	v	v	g	v
10	wadisa	34.00	1	260.00	v	g	g	d	g
11	swalih	32.00	2	400.00	v	v	a	g	a
12	swalih	58.00	1	310.00	d	v	g	v	v
13	swalih	21.00	2	200.00	g	d	v	g	g
14	abunas	28.00	1	400.00	d	g	b	g	b
15	khalda	58.00	2	270.00	v	v	a	v	a
16	jabilh	43.00	1	310.00	d	v	a	a	a
17	jabilh	22.00	2	230.00	v	a	b	v	g
18	jabilh	60.00	1	280.00	d	v	a	v	b
19	abunas	34.00	1	260.00	v	v	b	a	a
20	jubaih	40.00	2	220.00	g	g	a	v	a
21	swalih	19.00	1	330.00	d	g	a	v	g
22	khalda	51.00	2	550.00	a	v	b	g	g
23	jubaih	27.00	1	210.00	g	a	a	g	a
24	jabilh	38.00	1	320.00	v	a	b	g	g
25	swalih	31.00	1	190.00	g	g	a	v	g
26	abunas	45.00	2	320.00	b	v	a	g	a
27	wadisa	50.00	1	610.00	a	g	a	a	a

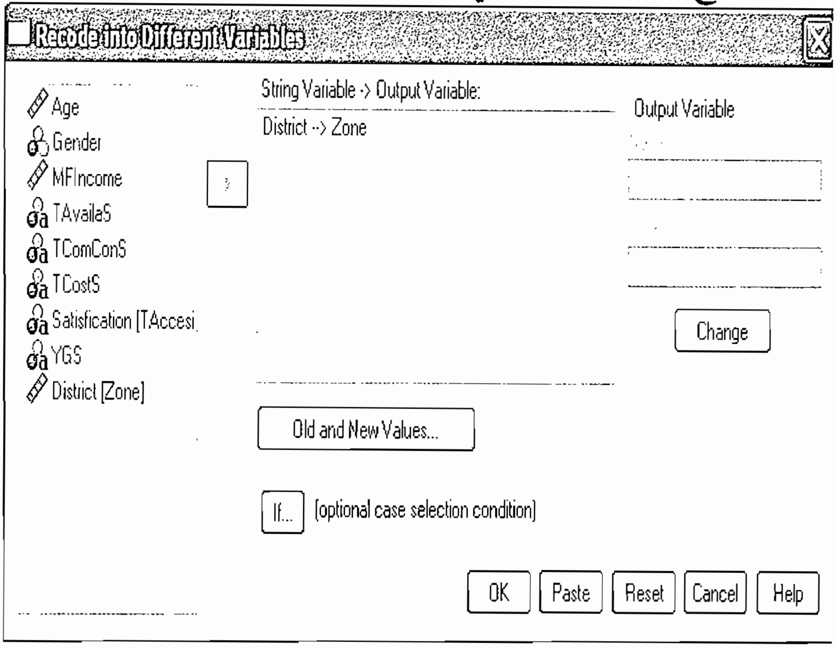
Data View | Variable View

SPSS Processor is ready

- اختيار الامر الفرعي Recode من قائمة Transform ومنه الى into different variable عندها سيتم فتح مربع الحوار المبين في الشكل رقم (4.1) ادناه، ويتم فيه نقل المتغيرالمطلوب تحويله وليكن متغير المنطقة District من قائمة المتغيرات الموجودة الى يسار مربع الحوار بواسطة ايقونة السهم ليصبح في اعلى المربع، ومن ثم تدوين رمزالمتغيرالجديد المطلوب تشكيله ولنرمز له بـZone مع امكانية تعريفه في خانة label، ثم النقر على ايقونة change .

شكل بياني رقم (4.1)

مربع حوار الامر الفرعي Recode من قائمة Transform



- النقرفوق ايقونة old and new value ليظهر لنا مربع الحوار التالي المبين في الشكل البياني رقم (5.1)، ويتم فيه التاشير على Old value ويدون تحتها رمز المنطقة كما جاء في الملف وليكن جبيهة Jubaih وتدوين 1 كرمز جديد تحت New value يلي ذلك الكبس على add ونستمر بتكرار الخطوة مع باقي رموز المتغيرولغاية الرمز khalda وتدوين الرمز الجديد بدله وهي القيمة 7 .

▪ الكبس على ايقونة Continue ومن ثم Ok ليظهر المتغير الجديد Zone برموزه الجديدة الكمية على صفحة ادخال المعطيات Data view تلقائيا ضمن الملف المعني.

▪ وبذات الاجراءات اعلاه تتم العملية مع باقي المتغيرات النوعية على لوحة الشكل البياني (5.1) وذلك باعادة تسمية كل منها لتصبح :

Avsat بدلا من TAvailS (مستوى الرضا عن توفر واسطة النقل)

Ccsat بدلا من TComConS (مستوى الرضا عن راحة وملائمة النقل)

Costsat بدلا من TCostS (متوى الرضا عن اجور النقل)

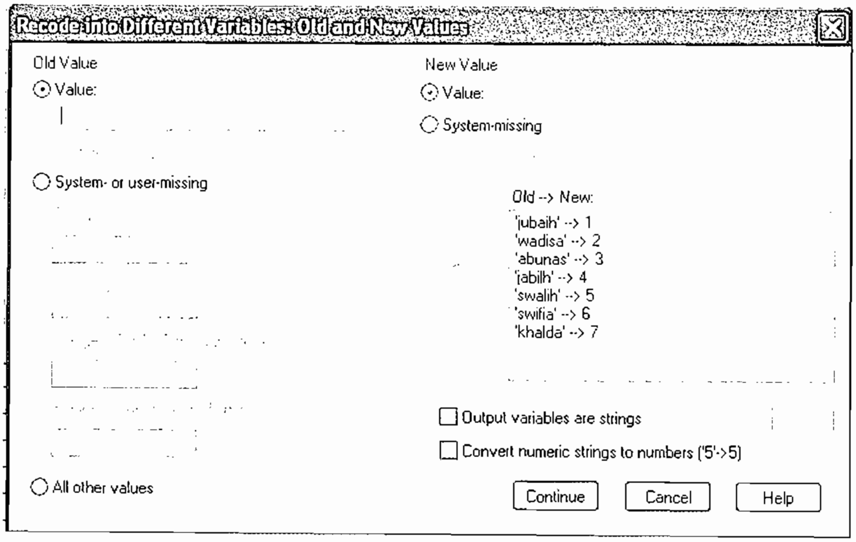
Accsat بدلا من TAccesbS (مستوى عن موقع توفر واسطة النقل)

Ysat بدلا من YGS (مستوى الرضا العام عن خدمة النقل العام)

واعطاء رموز جديدة لاجابتها وهي 1، 2، 3، 4، 5، على التوالي بدلا من : b (سيء)، a (مقبول)، g (جيد)، v (جيد جدا)، d (ممتاز)، فنحصل على جدول المعطيات بقيم كمية وكما مبين مقطع منه في الشكل البياني رقم (6.1).

شكل بياني رقم (5.1)

يوضح مربع الحوار التالي لتكملة ايعازات الامر الفرعي Recode



الشكل البياني رقم (6.1)

مقطع من مخرجات تحويل القيم النوعية الى كمية باستخدام الخيار Recode

Zone	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5.00	3.00	3.00	2.00	2.00	1.00
6.00	5.00	3.00	3.00	4.00	4.00
7.00	4.00	3.00	3.00	1.00	3.00
2.00	3.00	4.00	4.00	3.00	4.00
2.00	4.00	3.00	3.00	6.00	3.00
5.00	4.00	4.00	2.00	3.00	2.00
5.00	6.00	4.00	3.00	4.00	4.00
5.00	3.00	5.00	4.00	3.00	3.00
3.00	5.00	3.00	1.00	3.00	1.00
7.00	4.00	4.00	2.00	4.00	2.00
4.00	5.00	4.00	2.00	2.00	2.00
4.00	4.00	2.00	1.00	4.00	3.00
4.00	6.00	4.00	2.00	4.00	1.00
3.00	4.00	4.00	1.00	2.00	2.00
1.00	3.00	3.00	2.00	4.00	2.00
6.00	5.00	3.00	2.00	4.00	3.00
7.00	2.00	4.00	1.00	3.00	3.00
1.00	3.00	2.00	2.00	3.00	2.00
4.00	4.00	2.00	1.00	3.00	3.00
5.00	3.00	3.00	2.00	4.00	3.00
3.00	1.00	4.00	2.00	3.00	2.00
2.00	2.00	3.00	2.00	2.00	2.00

1- 6- 2 المنهجية التجميعية (Agreggate Methodology)

وهي التي تكون فيها المشاهدات (observations) التي تخضع لعملية التحليل معبرة عن مجموعة من المجتمع الاحصائي، وبذلك تكون وحدة المشاهدة على مستوى الاقليم او المحافظات او المناطق او الوحدات الزمنية وما شابه، وعليه فان المعطيات التي تستخدم عادة ما تكون متوسط Mean او معدل Ratio او نسبةPercentage، اي الاعتماد على قيمة مفردة لتكون ممثلة لمجموعة من القيم، والنماذج التي تستخدم مع هكذا نوع من المعطيات تدعى بالنماذج التجميعية.

وبذلك فان قيم المتغيرات التجميعية تمثل الخصائص المتصلة بهذه الاقاليم او المحافظات او المناطق او الزمن مما يجعل هذه النماذج تستجيب للمناطق الجغرافية التي تعود اليها المعطيات، بينما تكون استجابتها اقل

للتغيرات التي تطرأ على سلوك الافراد اتجاه السلعة او الخدمة، وعليه فهي محدودة المرونة عند استخدامها من قبل المنتج او متخذ القرار في السيطرة والتحكم بالاضافة الى محدودية فعالياتها عند تطبيقها على مناطق جغرافية اخرى من غير التي جمعت منها المعطيات. وبذلك فهي تصلح للحالات التي تكون فيها الدراسات شاملة او عامة كما هو الحال في دراسات النقل والمرور التي لابد من تقسيم منطقة الدراسة على مستوى مناطق المحافظة او على مستوى المحافظات او عند استخدام جداول نشرات المنظمات الدولية على اساس الاقاليم من خلال توزيع الدول على مجموعات كمجموعة الدول الصناعية ومجموعة الدول الاوربية ومجموعة الدول النامية وهكذا. وعليه فان هذا النوع من النماذج يكون ملائم للمشاريع التي يتوزع نشاطها على اقاليم مختلفة وكذلك ملائم للجهات المركزية الحكومية التي تقدم خدماتها على عموم البلد او في وضع الخطط التنموية الشاملة او على نطاق المحافظات او نطاق فروع شركة يتوزع نشاطها على مدن او دول عديدة.

ان تجميع المعطيات الواسع من اجل الحصول على متوسط او معدل او نسبة عن كل منطقة او اقليم من شانه ان يؤدي الى فقدان جزء كبير من المعطيات التي يتم جمعها وهذا من شانه حصول انخفاض كبير في عدد المشاهدات التي تخضع لعملية التحليل، وبالتالي انخفاض في درجة معنوية النتائج، حيث وكما هو معلوم كلما ازداد عدد المشاهدات (العينة) n على عدد المتغيرات الخاضعة للتحليل k كلما ازداد ضمان الحصول على معنوية اعلى. ومن بين الطرق الاحصائية المهمة التي تستخدم مع هذه النماذج هي تحليل الانحدار (regression analysis) ودالة التمييز (discriminite analysis) وتحليل المركبات الاساسية (principal component analysis).

ولاجل اعداد الجدول الملائم للتحليل لمعطيات المثال (1.1) بموجب المنهجية التجميعية agreggate methodology يتم توزيعهم على

المناطق (zones) وهو ما يستلزم القيام بايجاد متوسط الاجابات لكل منطقة، اي نحتاج لايجاد قيمة واحدة لكل منطقة لتكون ممثلة عن كافة الاستبانات التي تخص تلك المنطقة، وبذلك فان عدد المشاهدات سيكون مساويا لعدد المناطق المشمولة بالاستقصاء وهي 7 مشاهدات بدلا من 31 مشاهدة. فلو اخذنا على سبيل المثال متغيري العمر والدخل الشهري للاسرة واستخرجنا قيمها لمنطقة جبيهة، فسيكون لدينا الاتي :

منطقة جبيهة : عدد الاستبانات التي تخص هذه المنطقة هي 3 استبانة، فتصبح قيم معطيات المتغيرات المعنية كالاتي :

متغير العمر $\bar{x} = \frac{45 + 40 + +27}{3} = 37$

متغير الدخل الشهري للأسرة $\bar{x} = \frac{260 + 220 + 210}{3} = 230$

وبنفس الطريقة يتم استخراج معطيات باقي المناطق، في حين يتم ايجاد قيم المتغيرات النوعية الاخرى للاستقصاء بعد ان يتم تحويلها الى قيم كمية وفق الاسلوب الوارد في فقرة نماذج الراي من هذا الفصل وكما مبين في الجدول 1-2 ومن ثم ايجاد متوسطاتها ايضا، وعليه نحصل على جدول المدخلات للتحليل التجميعي المبين شكله في (1-3) ادناه :

جدول رقم (3-1) :

صيغة جدول المدخلات للمثال (1-1) للمنهجية التجميعية

Zone			Age	Sex	MFInc	avaisat	consat	costsat	accsat	ysat
جبيهة	1	Mean	37	1.33	230.0	2.67	3.67	1.33	3.00	2.33
وادي السير	2	Mean	32	1.20	292.0	3.20	3.20	2.60	2.80	2.80
ابونصير	3	Mean	36	1.50	328.3	3.50	3.83	1.67	3.50	2.50
جبل الحسين	4	Mean	42	1.20	290.0	4.40	3.60	1.80	3.20	2.20
صويلح	5	Mean	36	1.67	361.7	3.67	4.00	2.33	3.17	2.67
صويفية	6	Mea n	30	1.33	403.3	4.33	3.33	2.67	3.67	3.00
خلدا	7	Mean	48	1.67	380.0	3.00	3.33	2.00	2.67	2.67

1- 6- 3 حالة دراسية رقم C_{1-2}

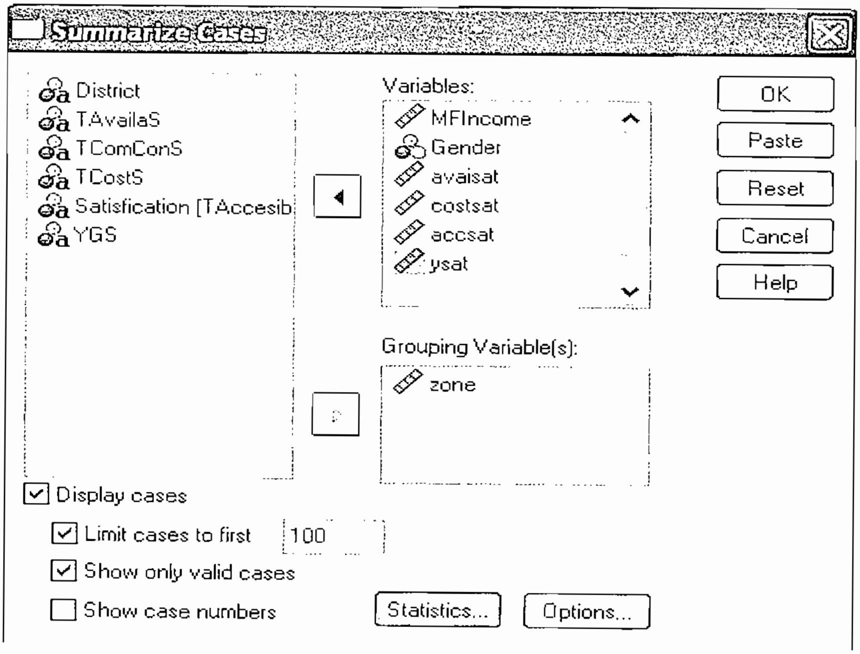
استخدام برنامج SPSS في الحصول على معطيات على مستوى المفردة والمستوى التجميعي

- استدعاء ملف معطيات المثال رقم (1.1) الذي تم انشاءه في اعلاه ،

- استخدام الخيارCase Summaries من الامر الفرعي Reports من قائمة Analysis، مع الاشارة الى ان حجم العينة N يجب ان لايزيد على 100 عند استخدام الامر الفرعي Reports ليظهر مربع الحوار المبين في الشكل البياني رقم (7.1)، ليتم فيه استخدام السهم الجانبي لنقل المتغيرات المطلوب اظهار قيمها وايجاد متوسطاتها الى تحت عنوان Varaibles، ونقل متغير Zone الذي تصنف بموجبه قيم المتغيرات الى تحت عنوان Grouping Varaible ،

الشكل البياني رقم (7.1)

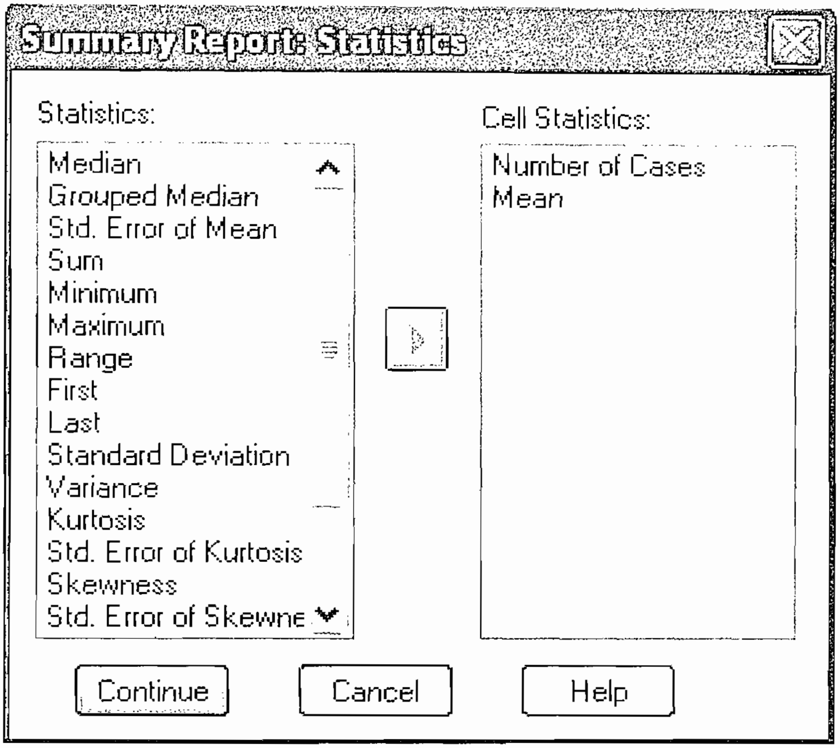
مربع حوار الخيارCase Summaries من الامر الفرعي Reports



▪ الكبس على ايقونة Statistics فتظهر لنا لوحة Summay Statistics : Report، لتظهر لنا اللوحة المبينة في الشكل البياني رقم (8.1)، ليتم فيها استخدام المقياس المطلوب توظيفه لتجميع المعطيات، وليكن المتوسط مثلا Mean، ثم الكبس على ايقونة Continue للعودة الى مربع الحوار،

الشكل رقم (8.1)

لوحة Summay Report : Statistics



▪ الكبس على ايقونة Ok فنحصل على جدول المخرجات رقم (4.1) الذي يضم قيم المتغيرات على المستويين المفردة Disagrggate والتجميعي Agreggate مصنفة حسب المناطق zones .

جدول رقم (4-1)

جدول مخرجات الخيارCase Summaries لمدخلات للمثال (1 -1)

	Zone	Age	Sex	MFI	TAv	TCC	TCos	TAcc	YGS
1	1	45	1	260	3	5	2	2	3
	2	40	2	220	3	3	1	4	2
	3	27	1	210	2	3	1	3	2
T	n	3	3	3	3	3	3	3	3
	Mean	37.33	1.33	230.00	2.67	3.67	1.33	3.00	2.33
2	1	25	2	180	4	4	2	1	3
	2	20	1	210	3	4	4	3	4
	3	34	1	260	4	3	3	5	3
	4	50	1	610	2	2	2	2	2
	5	31	1	200	3	3	2	3	2
T	n	5	5	5	5	5	5	5	5
	Mean	32.00	1.20	292.00	3.20	3.20	2.60	2.80	2.80
3	1	51	2	260	5	3	1	3	3
	2	19	2	500	3	5	3	4	5
	3	38	1	400	5	3	1	3	1
	4	34	1	260	4	4	2	3	2
	5	45	2	320	1	4	2	4	2
	6	27	1	230	3	4	1	4	2
T	n	6	6	6	6	6	6	6	6
	Mean	35.67	1.50	328.33	3.50	3.83	1.67	3.50	2.50
4	1	46	1	310	4	4	2	3	2
	2	43	1	310	5	4	2	2	2
	3	22	2	230	4	3	2	4	3
	4	60	1	280	5	4	1	4	1
	5	38	1	320	4	3	2	3	3
T	n	5	5	5	5	5	5	5	5
	Mean	41.80	1.20	290.00	4.40	3.60	1.80	3.20	2.20
5	1	52	2	420	4	4	2	2	1
	2	32	2	400	4	4	2	3	2
	3	58	1	310	5	4	3	4	4
	4	21	2	200	3	5	3	3	3
	5	31	1	190	3	3	1	3	3
	6	21	2	650	3	4	3	4	3

T	n	6	6	6	6	6	6	6	6
	Mean	35.83	1.67	361.67	3.67	4.00	2.33	3.17	2.67
6	1	22	1	450	5	3	3	4	3
	2	19	1	330	5	4	2	4	3
	3	48	2	430	3	3	3	3	3
T	n	3	3	3	3	3	3	3	3
	Mean	29.67	1.33	403.33	4.33	3.33	2.67	3.67	3.00
7	1	34	1	320	4	3	2	1	3
	2	58	2	270	4	4	2	4	2
	3	51	2	550	1	3	2	3	3
T	n	3	3	3	3	3	3	3	3
	Mean	47.67	1.67	380.00	3.00	3.33	2.00	2.67	2.67
T	Total n	31	31	31	31	31	31	31	31
	Mean	36.84	1.42	325.48	3.58	3.61	2.06	3.16	2.58

1- 7 صياغة المتغيرات Variables Formulation

وهي المعالجة التي يتم انجازها من اجل تحقيق واحد او اكثر من الاهداف التالية :

- تاهيل متغيرات هامة للظاهرة بحاجة لاعادة صياغة للتعبير عن علاقتها بالمتغير التابع،
- للتخلص من العلاقات المتداخلة بين متغيرات هامة بالنسبة للمتغير التابع ،
- تاهيل متغيرات لتاتي متماثلة في شكل الصياغة مع متغيرات اخرى تعود لجهة او منظمة اخرى من اجل اجراء مقارنة معها لقياس مثلا مستوى جودة المتغير المعني بالقياس لتلك الجهة او المنظمة ،
- تكييف المعطيات بما يتلائم واخضاعها لعملية التحليل الكمي، او لزيادة معنوية النتائج .

1- 7- 1 الصيغة المطلقة Absolut Form

وتعني ادخال القيمة الكمية المطلقة مباشرة في النموذج من دون ان تشتمل على مقارنة او من دون الربط بظاهرة او متغير اخر، وهذه الصيغة تقوم على اساس ان متخذ القرار او الزبون يقوم بصورة مباشرة بتطبيق قناعته بالسلعة او الخدمة المعنية من دون المقارنة بالخاصية المناظرة للسلع والخدمات الاخرى، وقد يعود سبب ذلك الى خبرته المسبقة او ادراكه من ان السلعة او الخدمة المعنية هي الافضل لاختياره.

1- 7- 2 الصيغة النسبية Relative Form

وهنا تعني من ان المقارنة هي نسبية بين الخاصية i مع افضل خاصية مناظرة j مثلا. ومغزاها ان الزبون يستعرض خاصية او اكثر وينسبها لافضل خدمة او سلعة منافسة. فاذا رمزنا للنسبة R ، وللسلعة او الخدمة بـ S فقيمة المتغير ستكون عبارة عن :

$$R = S_i / S_j$$

1- 7- 3 صيغة الفرق المطلق Absolute Difference

وهو ما يدعى بالتقييم وفقا للفرق المطلق، والذي يعبر عن الفرق في درجة القناعة بين خاصيتي خدمتين او سلعتين، فاذا رمزنا للصيغة AD فان قيمة المتغير هي عبارة عن :

$$AD = S_i - S_j$$

1- 7- 4 صيغة الفرق المطلق المرجح

Weighted Absolute Difference

وبموجب هذه الصيغة فان الفرق سينسب لعامل ما لاجل التخفيف من حدة الفرق المطلق، فاذا رمزنا للعامل الذي ينسب اليه الفرق F وليكن مثلا سهولة الوصول الى السلعة او الخدمة accessibility فان علاقة ايجاد قيمة المتغير تصبح :

$$WAD = \frac{S_i - S_j}{F}$$

1- 7- 5 صيغ اخرى

كالاستعانة بالمتغير التابع وادخاله في صياغة المتغير المستقل (بالقسمة او الضرب او الترجيح ...) وهو ما يطلق عليه Cross product او اخذ تربيع او تكعيب او لوغاريتم المتغير المستقل او التابع او لوغاريتم احد طرفي المعادلة Semi-Logarithems.... الخ

1- 8 أنواع المسوحات (الاستقصاءات) الاحصائية

1- 8- 1 المسوحات الشاملة Censuses والمعلمة Parameter

وهي المسوحات التي تشمل كافة مفردات المجتمع الاحصائي سواء كانت هذه المفردات (الوحدات) انسانا او نباتا او جمادا. كما هو الحال في المسوحات السكانية والصناعية والزراعية وغيرها، وهي ما يطلق عليها بالتعداد او الحصر الشامل.

ان هذا الاسلوب يحتاج الى امكانيات مالية وبشرية وفنية كبيرة ويحتاج ايضا لوقت طويل من التهيئة والتحضير. وغالبا ما يتم تنفيذ هذه المسوحات على فترات متباعدة نسبيا كان تكون كل 10 سنوات كما هو الحال في التعدادات السكانية والزراعية. ومن ابرز اهداف توفير معطيات كاملة عن المجتمع الاحصائي هي الاغراض الادارية الرسمية او لبناء خطط تنموية لاغراض اجتماعية واقتصادية، كما ويمكن الاستفادة من نتائج هذه المسوحات لاغراض تصميم العينات وفي تنفيذها باستخدامها كاطر احصائية

لاغراض سحب العينات وكادلة لاغراض التنفيذ. الا ان عدد هذه المسوحات اخذ بالتناقص تدريجيا في السنين الاخيرة نتيجة للعوامل التالية :

(1) التطور الكبير الحاصل في العمل الاداري للدول وانتظام السجلات الادارية والتوسع في استخدام الاجهزة الاكترونية.

(2) زيادة الوعي الاجتماعي والثقافي للافراد وادراكهم لاهمية اعطاء معطيات صحيحة عن اسرهم وممتلكاتهم وعناوينهم وغيرها ولحاجتهم اليها كمستمسكات رسمية في انجاز معاملاتهم عند الحاجة.

(3) تطور الاساليب العلمية الاحصائية والرياضية في مجال تعميم الاستنتاجات التي يتم الحصول عليها من العينات، وتيسير اساليب بناء التقديرات والتوقعات الدقيقة عن اجمالي المجتمع، وقد سهل ذلك وبدرجة كبيرة التوسع في استخدام الحاسوب الالي.

ويطلق على حصيلة مؤشرات المسوحات الشاملة بالمعلمة parameter والتي هي عبارة عن قيمة رقمية تصف خاصية معينة تعود للمجتمع الاحصائي (population) الذي يعني حجم المجموع، وعادة ما يعبر عن المعلمة بحرف لاتيني كبير، فمثلا معلمة الوسط الحسابي للمجتمع يعبر عنه بالحرف μ (ميو) والانحراف المعياري σ (سكما) وهكذا.

1- 8- 2 مسوحات العينة

Sampling surveys والاحصاءة Statistic

ان المسح بالعينة يعني شمول جزءا من المجتمع الاحصائي، على ان يكون هذا الجزء ممثلا دقيقا لخصائص المجتمع. ومن الامثلة على هذا النوع من المسوحات استطلاعات الراي ومسوحات الاسرة وخدمات النقل والخدمات

الاجتماعية والاقتصادية والظواهر الحياتية وغيرها. ومن اهم ميزات اسلوب العينات هي :

(1) توفير الوقت والجهد والتكاليف.

(2) توقع الحصول على نتائج المسح بوقت قصير.

(3) زيادة دقة المعطيات الاحصائية نتيجة لقلة الاخطاء البشرية التي تشكل بحدود 90٪ من إجمالي اخطاء المسوحات وذلك كنتيجة لاستخدام عدد قليل من الايدي العاملة مقارنة لما تحتاجه المسوحات الشاملة.

(4) توفر الطرق العلمية المناسبة للعينات كمقياس فترة الثقة confidence limits واختبار الفروض hypothesis testing وغيرها التي تتيح الفرصة للتاكد من مستوى دقة نتائج مسوحات العينة.

(5) هناك حالات استحالة لاستخدام المسوحات الشاملة كما هو الحال مع المجتمعات الانهائية كالاسماك والطيور وما شابه، وكذلك مع الحالات التي تؤدي لخسائر كبيرة او تتسبب بتكلفة باهضة اذا ما اجري المسح الشامل عليها في الانتاج والطب والمواد الغذائية وغيرها، مما تستوجب استخدام العينات معها حصرا.

اما المؤشرات الناتجة عن العينة فتدعى **الاحصاءة statistic**، اي انها قيمة رقمية تصف خاصية معينة تعود لعينة (sample)، ويعبر عن الاحصاءة بحروف انكليزية، فاذا كنا بصدد الوسط الحسابي للعينة نرمز لها $\bar{x}$ وللانحراف المعياري نرمز له s وهكذا.

1- 9 أطار المجتمع الاحصائي

وهو عبارة عن وصف لما هو متوفر من معطيات عن وحدات (مفردات) المجتمع الاحصائي المطلوب دراسته، ويعتبر الاطار احد المستلزمات الاساسية لتصميم العينة سواء عند تحديد نوع العينة اوعند تحديد حجمها اوعند سحب وحداتها. كما ويعتمد الاطار ايضا في تنظيم العمل الميداني ولاغراض ادارة المسح او الاستقصاء وعند تنفيذه من خلال ماتوفره من معطيات عن مواقع وعناوين وحدات العينة المشمولة بالمسح.

وقد تتخذ الاطر شكل خارطة تضم المواقع المطلوب بحثها كالمقاطعات او القرى او المزارع، او شكل قوائم تضم اسماء وعناوين وحدات المجتمع الاحصائي، فاذا كانت الوحدة الاحصائية هي المسكن مثلا، فان الاطار يصبح عبارة عن قائمة تضم ارقام المساكن وعناوينها، واذا كانت الدراسة تخص المستشفيات مثلا، فيكون الاطار عبارة عن قائمة شاملة لاسماء وعناوين المستشفيات واختصاصاتها وعدد الاسرة فيها وهكذا. وتعتبر نتائج المسوحات الشاملة هي المصدر الاساس في توفير الاطر الاحصائية، وفي كثير من الحالات يتم الاعتماد ايضا على قوائم الماء والكهرباء المتوفرة لدى المؤسسات المعنية في توفير الاطر الاحصائية. ولكي يكون الاطار صالحا ينبغي ان يكون حديثا ويعود لفترة زمنية قريبة، وان يضم جميع وحدات المجتمع المستهدف دراسته، وان لايكون هناك تداخل بين وحداته (اي عدم تكرار ظهور اي من الوحدات).

الفصل الثاني

تصميم الاستبيان (الاستمارة) الاحصائية

Questionnaire Design

2-1 التغطية او الشمول Coverage

ويقصد بها تحديد طبيعة وحجم المعطيات المطلوب جمعها بواسطة الاستبانة الاحصائية المستهدف توفيرها للبحث والاماكن التي ستجمع منها المعطيات وتفاصيلها، وهو ما يطلق عليه بالتغطية Coverage، ويعود ذلك لضرورة تحديد الامكانيات المطلوبة للبحث اوفي مجال المتابعة الميدانية عند تنفيذ المسح بالاضافة للمبررات الاتي ذكرها.

2-1-1 التغطية المكانية Spatial Coverage

ويقصد بها المساحة الجغرافية التي سيشملها البحث. فهي تساعدنا من معرفة ان كانت المناطق المشمولة هي مناطق بادية او ريف او حضر، حيث لكل منها خصائص تعليمية وثقافية وتقاليد ومفاهيم لمفردات لغوية تستوجب مراعاتها عند صياغة اسئلة الاستبانة وفي تحديد طريقة ملئ الاستبانات، فتكون بطريقة المقابلة المباشرة في حالة الريف او البادية لاحتمال انتشار الامية او بطريقة التدوين الذاتي ان كانت في الحضر لارتفاع نسبة المتعلمين والى غير ذلك، كما ان تحديد التغطية المكانية يساعدنا ايضا في معرفة مدى الشمول ان كان على نطاق الدولة، ام على مستوى اقليم او محافظة معينة، فلو كان على مستوى المحافظة مثلا، فلابد من معرفة ان كان المقصود هو شمول كافة مناطق المحافظة ام مناطق معينة منها، لان لكل منطقة خصائص

قد تختلف عن اخرى مما يستدعي مراعاتها في طبيعة اسئلة الاستبانة اوفي ادارة المسح ومتابعته .

2-1-2 التغطية النوعية (المعطيات) Data Coverage

ويراد به معرفة ان كان البحث سيشمل كافة اوجه النشاط المعني ام جزء منه، وما هو هذا الجزء. فلنفترض ان البحث يخص قطاع النقل مثلا فيستوجب التحديد، هل نعني بذلك نقل الركاب ونقل البضائع، واذا كان المقصود هو نقل الركاب فقط فهل يشمل النقل العام والخاص ام احدهما واذا كان النقل العام هو المعني، فهل يشمل نقل الركاب داخل المدن فقط ام يشمل النقل بين المدن ايضا وهكذا، ليتم في ضوء ذلك تحديد طبيعة الاسئلة المطلوبة .

2-2 القواعد العامة لتصميم الاستبانة

2-2-1 **ان يكون حجمها مناسبا**، وان يكون نوع الورق غير قابل لانتشار الكتابة، ويكون لونه مقبولا، وتكون الطباعة جيدة وسهلة القراءة، واذا كانت الاستبانة مكونة من عدة صفحات فانه يستحسن ان تكون على شكل كراس .

2-2-2 **مراعاة التنفيذ الالي لتبويب** وتحليل المعطيات وذلك بتخصيص حقول للرموز coding وتكييف الاستبانة بما تلائم وهذا الغرض، مع مراعاة عملية الترميز المسبق pre-coded اقتصادا في الجهد والوقت وكما مبين في نموذج الاستبانة التي سنتطرق اليها.

2-2-3 **ان تضم الحد الامثل من الاسئلة** والتاكد من خلوها من الاسئلة التي لاتخدم اهداف البحث او الدراسة .

4-2-2 مراعاة المفاهيم والتصانيف الاحصائية الدولية statistical definitions and classifications بهدف تحقيق التوحيد في مفهوم definition البيان او المعلومة وموقعها ليتسنى اجراء المقارنات الدولية والاقليمية والزمنية، كما ان اعتماد هذه الادلة يساعد في توفير الوقت والجهد في عملية الترميز وفي تبويب المعطيات. خاصة وان هذه الادلة معتمدة رسميا من قبل كافة الدول الاعضاء في الامم المتحدة وتقوم المنظمات والمؤسسات الدولية المختصة باعدادها والتوصية باستخدامها وتتكييفها وفقا لظروف وخصائص كل بلد، ومن هذه الادلة دليل النشاط الاقتصادي القياسي الدولي ISIC والتصنيف القياسي الدولي للتجارة SITC والتصنيف القياسي الدولي للمهن ISCO والتصنيف القياسي الدولي للتعليم ISCE وغيرها. والملاحق من رقم (1.2) الى (3.2) تبين مقاطع كنماذج لبعض من هذه الادلة والتصانيف .

3-2 شروط صياغة أسئلة الاستبانة

1-3-2 الاخذ بنظر الاعتبار ان الاسئلة موجهه الى افراد قد يختلفون في مستوياتهم ومؤهلاتهم الثقافية والتعليمية وحتى في عاداتهم الاجتماعية، مما يستدعي الوضوح في صياغة الاسئلة من خلال استخدام عبارات بسيطة ولها معنى مالوف وتعطي في ذات الوقت المعنى المقصود. فمثلا هناك اختلاف في مفهوم الشراكة عند سكان المدن عنه في البادية، ففي المدن يرتبط المفهوم براسمال لمزاولة نشاط تجاري او اقتصادي، في حين يرتبط مفهوم الشراكة في البادية بالمشاركة في قطيع الماشية من جمال او اغنام وفي حصص الانتاج الحيواني من هذه الماشية. ويطلق على الشراكة في المدينة بالمساهمة و التضامنية وذات المسؤلية المحدودة ...... الخ، في حين يطلق على الشراكة في البادية بشراكة السمونة وشراكة العظم .... الخ .

2-3-2 ضرورة تجنب الاسئلة الغامضة، كأن يسأل المسافر مثلا عما اذا كان مستوى النقل هذا العام افضل من مستوى النقل في العام الماضي، مما يجعل الاجابة صعبة ومعقدة لعدم توضيح معنى المعيار ان كان المقصود دقة المواعيد ام راحة السفر ام سرعته ام كلفته ام توفره وما الى ذلك.

3-3-2 ان تصاغ الاسئلة لتكون الاجابة عنها قاطعة وتاتي على شكل رقم او كلمة نعم او لا او بوضع اشارة معينة وهكذا، كما ويفضل حصر وتحديد الاجابات المحتملة عن كل سؤال وكتابتها امام السؤال المعني ليقوم المبحوث respondent بوضع الرقم او الاشاره او غيرها على الاجابة المناسبة تسهيلا لتسجيل الاجابة واختصارا في الوقت والجهد مع تحقيق هدف التوحيد في المعنى. فعند السؤال عن المستوى التعليمي للمبحوث مثلا، تحدد المستويات التعليمية المتوقعة امام السؤال ليقوم المبحوث بوضع الاشارة عند المستوى التعليمي الذي هو حاصل عليه.

4-3-2 ضرورة ترتيب الاسئلة ترتيبا منطقيا يراعى العلاقة فيما بينها، وتقسم الى مجموعات متجانسة تحمل عناوين فرعية، وعلى ان نبدأ بالاسئلة السهلة الاجابة التي لاتحتاج الى تفكير كتلك المتعلقة بخصائص الشخص مثلا كالاسم والجنس والعمر والعنوان والتحصيل الدراسي وهكذا.

5-3-2 ان لا تكون الاسئلة من النوع الايحائي، اي التي توحي باجابات معينة، فلا يسأل مثلا هل انت متدين؟ لانه ليس من المنتظر ان تاتي الاجابة بالنفي، ولكن بالسؤال مثلا عما اذا كان يؤدي بعض الشعائر الدينية .

6-3-2 ان تكون الاسئلة قدر الامكان بعيدة عن الحساسية او الاحراج وان لا تعد تدخلا في مسائل شخصية تؤدي الى ازعاج المبحوث ورفضه للتعاون.

7-3-2 الا تكون الاسئلة من النوع الذي يقود للتحيز، فلا يسال مثلا هل تاخرت بسبب النقل او ازدحام المرور؟ لان الاجابة ستكون في الغالب بنعم، لان الاشارة الى السبب تتضمن دلاله على الاتفاق عليه وان لم يكون هو السبب الرئيسي للتاخير او عدمه، ولا يسال مثلا هل تشتري الصحف يوميا؟ فقد يدفع حب التفاخر الى الادعاء بشرائها، في حين بالامكان ان يتضمن السؤال ايضا هل تطلع على الصحف وهكذا.

8-3-2 عدم تضمين الاسئلة اكثر من نقطة واحدة، فاذا كان للسؤال جزءان يستحسن ان يكونا سؤلين متتالين، فلا يسال مثلا هل تمتلك فيديو وتلفزيون، او هل تستمع للاذاعة والتلفزيون، فمن الجائز ان يمتلك احدهما فقط في السؤال الاول او يستمع لاحدهما في الثاني. وفي حالة اللجوء الى دمج سؤالين معا مثل هل تشاهد التلفزيون واي البرامج تفضل، ينبغي اللجوء الى طريقة الاسئلة ذات الابعاد المتعددة multidimentional scalling التي تم التطرق اليها في الفصل الاول، ليصبح السؤال كالاتي : اذا كانت الاجابة بنعم على مشاهدة التلفاز فاي البرامج تفضل .

9-3-2 يجب ان لاتكون الاسئلة من النوع المفتوح الذي يسمح باحتمال تعدد الاجابة عنها، مما يتطلب تحديد نوع الاجابات مسبقا، فعند السؤال عن الهواية فمن المفضل ان تكتب امام السؤال عددا من الهوايات الرئيسية ليؤشر البحوث احدها، وان لم تكن هوايته من بين المثبتة امام السؤال فانها تدخل ضمن فقرة اخرى الذي ينبغي توفرها.

10-3-2 ان تاتي صياغة الاسئلة بحيث لا تتطلب من المبحوث اجراء عمليات حسابية مطولة او تستدعي ذاكرة حادة ومجهودا فكريا، فلا يسأل مثلا كم هو عمرك في تاريخ معين بل يكتفى بالسؤال عن تاريخ ميلاده

ليقوم الباحث بعد ذلك باجراء عملية الطرح لمعرفة العمر، او يسال مثلا عن معدل الافراد في الغرفة الواحدة للمسكن بل يكتفي بالسؤال عن عدد افراد الاسرة وعن عدد الغرف للتوصل لذلك .

2-3-11 ضرورة ذكر الوحدات القياسية مثل العدد، كيلو، قدم....الخ مع تفضيل استخدام المقاييس الكمية والابتعاد عن المقاييس الكيفية التي تعتمد على تقدير الاشخاص، فلا يسال مثلا هل تذهب الى المكتبة العامة كثيرا، بل يستحسن تحديد عدد المرات ليصبح السؤال اذكر عدد المرات التي تزور فيها المكتبة اسبوعيا مع تحديد عدد المرات على شكل فئات مثل: 0-2، 3-5، 6 فاكثر ليقوم المبحوث بتاشير احدها.

2-3-12 من المفضل اضافة بعض الاسئلة بصيغ مختلفة لابقصد الاجابة عليها لذاتها وانما للتاكد من دقة الاجابات الاخرى، كان يسال في البداية عن متوسط دخل الاسرة الشهري وفي موقع لاحق يسال عن متوسط مصروف الاسرة الشهري ليتم المقارنة بين الاجابتين .

2-4 الاجزاء التي تتكون منها الاستبانة

عندما يتولى الباحث اجراء المقابلة وشرح اسئلة الاستبيان وتدوين المعطيات بنفسه فان الاستبيان سيتكون من جزئين فقط هما الجزء الاول والثاني اللذان سيلي شرحهما، اما في الحالة التي يقوم المبحوث respondent الاجابة على الاسئلة وتدوين هذه الاجابات فستتكون الاستبانة من ثلاثة اجزاء وكما هو مبين في نموذج الاستبانة عند نهاية الفصل، والاجزاء هي:

2-4-1 الجزء الاول: ويتضمن المعلومات المتعلقة باسم الجهة القائمة بالبجث وعنوانها وذلك في الجانب الايمن من اعلى الاستبانة، يليه في وسط

السطر عنوان البحث او موضوعه، يلي ذلك مقدمة تعريفية مختصرة بهدف البحث واهميته مع الاشارة في هذه المقدمة ايضا الى ان المعطيات التي سيدلي بها ستكون سرية وللاغراض العلمية حصرا مع الاشادة بتعاونه في انجاز البحث .

2-4-2 **الجزء الثاني**: ويتم في هذا الجزء ترتيب الاسئلة ويكون لكل مجموعة منها عنوان فرعي يدل على طبيعة مجموعة الاسئلة، ويراعى في الترتيب وكما اشرنا موضوع التجانس ومنطقية التسلسل، مبتدئين بالاسئلة البسيطة التي لاتحتاج الى تفكير كالاسم والعنوان والعمر وما شابه ومنتهين بالاسئلة الاكثر حاجة الى تفكير .

3-4-2 **الجزء الثالث**: ويكون هذا الجزء خاص بالتعليمات المتعلقة بشرح الاسئلة وتفسيرها، وكيفية ملئ الاستبانة، وكل ما من شانه جعل الاستبانة والاسئلة واضحة ومفهومة. وبصورة عامة يفضل ان تكون التعليمات على شكل كراس منفصل اذا كانت التفاصيل واسعة المضامين والاستبانة بحاجة لشرح مطول. مع التاكيد على مراعاة المفاهيم والتصانيف القياسية الدولية التي سبق الاشارة اليها في وضع التفصيل اذا تطلب الامر ذلك .

5-2 حالات دراسية في تصميم استبانات

1-5-2 حالة دراسية C_{2-1}

تصميم نموذج لأستبانة من ثلاثة اجزاء

احدى المؤسسات التي يتوزع نشاطها على فروع صناعية واعمال انشائية وزراعية مختلفة، تنوي القيام بتوفير معطيات عن العاملين لديها واستطلاع اراءهم بشان مدى رضاهم عن مستوى الخدمات التي تقدم لهم، والمطلوب تصميم استبانة يتم بواسطتها جمع المعطيات التي تغطي الاهداف التالية:

1. خصائص العمال العاملين لديها من ناحية مستوى مهاراتهم وحالتهم التعليمية واعمارهم وجنسهم في كل من الانشطة التي تقوم بها المؤسسة،
2. الوقوف على ارائهم ومستوى رضاهم عن الخدمات التي تقدمها المؤسسة في مجال التغذية وظروف العمل الصحية والطبية وعلاقات العمل،
3. توزيع نتائج المسح حسب:
   - نوع النشاط الاقتصادي للشركة
   - المستوى التعليمي للعامل
   - مستوى مهارة العامل
   - الفئات العمرية
   - الجنس

إجراءات تصميم الاستبانة :

- من خلال الاطلاع على اهداف المسح الاحصائي نجد نحن بحاجة الى ثلاثة مجاميع من الاسئله لتغطية هذه الاهداف متمثلة بـ: طبيعة النشاط الاقتصادي الذي يقوم به العامل، وخصائصه (العمر– الجنس– الحالة التعليمية– مستوى المهارة)، ورأيه عن مستوى الخدمات التي تقدمها له المؤسسة .
- يتم الاستعانة بالتصنيف القياسي للانشطة الاقتصادية ISIC واعتماد تسلسله كترميز codes لهذه الانشطة .
- سنفترض بان الاستبانة سيتم ملئها من قبل المبحوثين وذلك اقتصادا في الكلفة، وعلية ستتضن الاستبانة الجزء الثالث المتعلق بالتعليمات .

وعليه يصبح الشكل العام للاستبانة كالاتي:

المؤسسة العامة للانشطة الاقتصادية

تطوير الخدمات المقدمة للعاملين في المؤسسة

اخي العامل / اختي العاملة

ان الغرض من هذا البحث هو لتحسين الظروف الاقتصادية والخدمية للعاملين في شركات المؤسسة وان تعاونك في اعطاءكم المعطيات الصحيحة يساهم في تحقيق هذا الهدف، علما بان المعطيات ستستخدم حصرا للاغراض العلمية، ولاحاجة لذكر الاسم. وشكرا لتعاونكم

القسم الاول: موقع العمل وطبيعته	الترميز
1. اسم المدينة..	☐
2. اسم الشركة التي تعمل بها..	☐
3. طبيعة النشاط الاقتصادي للعمل الذي تقوم به	☐
- زراعة وصيد وغابات وصيد الاسماك 01	
- صناعات استخراجية 02	
- صناعات تحويلية 03	
- كهرباء وماء وغاز 04	
- تشييد وبناء 05	
- تجارة جملة ومفرد ومطاعم وفنادق 06	
- نقل ومواصلات وتخزين 07	
- تامين وتمويل وعقارات وخدمات مقدمة لقطاع الاعمال... 08	
- خدمات المجتمع والخدمات الاجتماعية والشخصية 09	
القسم الثاني : خصائص العامل	
1. الجنس : ذكر [ ] انثى [ ]..........	☐
2. العمر: 18-29 سنة [ ] 30-50 [ ] 51 فاكثر [ ].......	☐
3. الحالة التعليمية : امي [ ] يقوأ ويكتب [ ] ابتدائيه [ ]...	☐
ثانوية [ ] اعدادية [ ] كلية [ ] جامعة [ ]	☐
4. درجة المهارة : ماهر [ ] شبه ماهر [ ] غير ماهر [ ]....	☐
القسم الثالث : مستوى الرضا عن الخدمات التي تقدمها المؤسسة	

يرجى وضع علامة √ امام الاجابة المناسبة لكل من الخدمات التالية :	ممتاز	جيد جدا	جيد	متوسط	ردئ	
1. خدمات المطعم	( )	( )	( )	( )	( )	□
2. الخدمات الطبية	( )	( )	( )	( )	( )	□
3. ظروف العمل	( )	( )	( )	( )	( )	□
4. العلاقات العامة	( )	( )	( )	( )	( )	□
5. خدمات النقل	( )	( )	( )	( )	( )	□

تعليمات ملئ الاستبانة

- **القسم الاول من الاستبانة : موقع العمل وطبيعته**

السؤال 1 : يدرج اسم المدينة التي يقع فيها موقع عملك الحالي

السؤال 2: يدرج اسم المصنع او المعمل او الشركة لعملك الحالي

السؤال 3: يرجى وضع الرقم المبين امام طبيعة النشاط الذي تقوم به حاليا في المربع الثالث من حقل الترميز، مع الاشارة الى انك معني بالنشاط الذي تمارسه فعليا، اي اذا كنت تعمل في المطعم فتدون الرقم 06 الخاص بنشاط المطاعم والفنادق رغم انك تابع لمعمل نشاطه الرئيسي هو صناعة، ولو كنت تعمل زراعي في حدائق المصنع تقوم بتدوين الرقم 02 الخاص بنشاط الزراعة وهكذا. وذلك وفقا للمفاهيم القياسية الدولية للامم المتحدة.

- **القسم الثاني من الاستمارة : خصائص العامل**

يرجى فقط وضع علامة X في الحقل المناسب لاجابتك لكل من الاسئلة الواردة في هذا القسم من الاستبانة من دون الحاجة الى استخدام مربع الترميز. حيث سيقوم الباحث المختص بتحويل اجابتك الى قيمة رقمية في كل مربع في حقل الترميز و حسب علاقته بالاجابة .

- **القسم الثالث من الاستمارة : درجة الرضا عن مستوى الخدمات**

ايضا المطلوب في هذا القسم هو وضع علامة X تحت مستوى

الرضا الذي تعتقد به ان كان ممتاز او جيد جدا او جيد او متوسط او ردئ ولكل نوع من الخدمات الواردة في كل سؤال، ليقوم الباحث ايضا بعد ذلك بتحويل اجابتك الى قيمة رقمية في المربع المختص في حقل الترميز.

2-5-2 حالة دراسية C_{2-2}

تصميم نموذج لأستبانة من تتكون من جزئين

المطلوب تصميم استبانة لدراسة تستهدف الوقوف على اراء عينة من التدريسيين في الجامعات بشان الانتاج البحثي، وتحديد العوامل المؤثرة على هذا الانتاج من خلال تحديد درجة رضاهم عن ظروف وبيئة العمل البحثي الحالية، ومعرفة خصائص المبحوثين .

إجراءات تصميم الاستبانة :

عند الاخذ بنظر الاعتبار طبيعة المجتمع الاحصائي المبحوث من ناحية مستواهم التعليمي الذي يمكنهم من فهم الاستلة وتدوين اجاباتهم، مع وضوح الهدف والاسئلة، فان الاستبانة لاتحتاج الى تعليمات لشرحها، وبذلك تتكون من جزئين، لتاخذ الشكل التالي :

المركز العربي للبحوث والدراسات
الاردن – عمان

' استبانة احصائية '

دراسة مستوى رضا الباحثين

والعوامل المؤثرة على العمل البحثي في الجامعات العربية

الاستاذ/ الاستاذة المحترمين

يستهدف البحث الوقوف على عوامل تشجيع العمل البحثي وتشخيص معوقاته في وطننا العربي للاخذ بسبل دفعه وتطويره من اجل الرقي بالمجتمع وتنميته، وان اسهامك بالراي الصحيح في ملئ هذه الاستبانة هو السبيل لتحقيق هذا الهدف، مشيرين الى عدم الحاجة الى ذكر الاسم ومن ان المعلومات التي تدلي بها هي للاغراض العلمية حصرا .

شاكرين لكم تعاونكم واسهامكم

القسم الاول : معلومات عامة

خاص بالحاسوب

1. الجنس : ☐ انثى ☐ ذكر
2. العمر : (سنة)
3. التحصيل العلمي : ☐ دكتوراه ☐ ماجستير ☐ بكلوريوس
4. دولة الحصول على اخر شهادة ☐ بريطانيا ☐ امريكا ☐ دولة عربية ☐ اخرى

اللقب الاكاديمي:☐ استاذ ☐ استاذ مشارك ☐ استاذ مساعد ☐ مدرس ☐ مدرس مساعد

5. فترة الخدمة الوظيفية : في المؤسسات الاكاديمية (سنة)

   غير الاكاديمية (سنة)
6. الجامعة او المؤسسة الاكاديمية التي تعمل بها حاليا : ....
7. معدل الدخل الشخصي الشهري مقربا بالدولار: (دولار)
8. معدل الدخل الشهري للاسرة شهريا مقربا بالدولار: (دولار)
9. الاختصاص : ☐ علوم طبيعية ☐ اداب وشريعة ☐ ادارة وتجارة ☐ علوم طبية ☐ علوم هندسية ☐حاسوب وتكنولوجيا

القسم الثاني : مستوى الرضا على النطاق الشخصي

السؤال	غير كافي (1)	مقبول (2)	جيد (3)	جيدجدا (4)	عالي (5)
1. الدخل الشهري					
2. اجواء العمل المكانية (مستلزمات الراحة)					
3. توفير المستلزمات البحثية (اجهزة-دوريات)					
4. العلاقات العامة مع المسئولين					
5. الحصول على حقوق الترقيات العلمية					
6. المساواة في مزايا المكافئات والايفاد والتعامل					
7. الموازنة بين عدد المحاضرات والعمل البحثي					
8. الاهتمام بالكفاءة العلمية					
9. الاهتمام بالمبادرات والمقترحات التطويرية					
10. المحفزات المالية والمعنوية					

القسم الثالث : مستوى الرضا العام

السؤال	ضعيف (1)	مقبول (2)	جيد (3)	جيدجدا (4)	عالي (5)
1. جدية المؤسسة التي تعمل بها بتشجيع البحوث					
2. التعليمات والقوانين النافذة على البحوث					
3. تعاون المؤسسات الاخرى في توفيرالمعلومات					
4. جدية الجهات المستفيدة في تطبيق نتائج البحوث					
5. رعاية وكفاية المجلات التي تقوم بنشر البحوث					

القسم الرابع : الانتاج العلمي

1. عدد المؤلفات المنشورة : (كتاب)
2. عدد البحوث المنشورة : (بحث)

القسم الخامس : تعليقك ومقترحاتك ان رغبت

....

....

2- 6 طرق جمع المعطيات Methods of Data Collection

كنتيجة لاختلاف طبيعة المجتمعات الاحصائية واختلاف المعطيات التي نحتاج جمعها مع تباين الظروف والامكانات المالية المتاحة للبحث، فقد تعددت طرق جمع المعطيات ايضا، وبصورة عامة هناك خمس طرق رئيسية شائعة الاستخدام في اغلب الاستقصاءات والمسوحات الاحصائية، نستعرض في الاتي مفهوم والحالات الملائمة لاستخدام كل منها .

2- 6- 1 طريقة المشاهدة Observation Method

وهي الطريقة التي بموجبها يكون جمع المعطيات عن طريق مراقبة الظواهر كما هي على الطبيعة، وتستخدم في حالتين هما :

(1) مراقبة الظواهر مع استخدام المنطق في تفسير مايقع، وتستحدم عادة في بعض الحقول العلمية كالدراسات الاجتماعية والنفسية والتربوية والمختبرية. ومن امثلة ذلك معايشة الباحث لبعض فئات المجتمع لمراقبة نمط حياة هذه الفئات وما يحصل لاعضائها خلال تعاملهم ومناقشاتهم، كما هو الحال عند دراسة مجتمع السجناء او مجتمع البادية او قيام باحث باجراء تجارب في مختبر علمي ومرقبة النتائج وتدوينها وما شابه .

(2) مراقبة الظواهر لغرض التدوين (التسجيل) فقط، وكما يحصل عند وقوف الباحث مثلا عند نقطة معينة لتسجيل حركة المرور ونمطها، متمثلة بعدد وسائط النقل وانواعها واتجاهها عند نقطة معينة.

مزايا وعيوب طريقة المشاهدة :

كما يتضح فان القائمين باستخدام الحالة الاولى هم من الكوادر المؤهلة والمدربة ولها خبرة جيدة في مجال عملها، وبذلك فمن المتوقع ان تقل

الاخطاء مع استخدامها، كالاخطاء الناتجة عن غموض الهدف او عدم وضوح مفاهيم المعطيات، بالاضافة الى اختفاء اخطاء عدم الاستجابة .

اما عيوب الطريقة فتنحصر بكلفتها المرتفعة لحاجتها لكوادر مؤهلة خاصة مع الحالة الاولى .

2- 6- 2 طريقة التسجيل الذاتي Self-Recording Method

وتعني قيام الاشخاص المبحوثين بتدوين اجاباتهم بانفسهم على الاسئلة الواردة في الاستبانة. وتعد الطريقة فعالة في الحالة التي يكون موضوع الاستقصاء يهم المبحوثين مباشرة، كما في حالة الاستفسار عن طبيعة السكن الذي يرغبون فيه او لغرض شمولهم باعفاءات ضريبية او تقديم خدمات مجانية او مخفضة وماشابه. وتاخذ الطريقة واحد او اكثر من الاساليب التالية:

(1) يقوم الباحثون بزيارة وحدات المجتمع المشمول وشرح هدف الاستقصاء واهميته، ومن ثم ترك الاستبيان معهم ليقوموا بملئها في وقت لاحق، ويتم الاتفاق على موعد عودة الباحثين للقيام بجمعها بعد اتمام ملئها، وتساعد هذه الطريقة على التاكد ميدانيا من ملئ الاستبيانات بشكل صحيح ودقيق.

(2) ترسل الاستبيانات بالبريد الى المبحوثين للقيام بملئها، ثم يتم جمعها في وقت لاحق من قبل الباحثين، او الطلب الى المبحوثين اذا كانوا يعملون في مؤسسات او دوائر رسمية باعادتها بالبريد ايضا .

ميزات وعيوب طريقة التسجيل الذاتي :

تمتاز الطريقة بانخفاض كلفتها وخاصة عند الاعتماد على البريد، وبتحاشي تحيز الباحثين، كما انها تتيح الوقت الكافي للمبحوثين للاجابة على اسئلة الاستبيان، وتظهر فائدتها بشكل اكبرعند استخدام البريد اذا كانت وحدات العينة متباعدة وموزعة على مناطق متعددة .

اما عيوب الطريقة فتبرز عند وجود نسبة من المبحوثين لايهتمون باعادة الاستبيان، اما بسبب الكسل او عدم الرغبة باعطاء معلومات، او اعادتها ناقصة لعدم فهم بعض الاسئلة.

2- 6- 3 طريقة المقابلة الشخصية Interviewing Method

وفيها يتم جمع المعطيات عن طريق اتصال الباحث شخصيا بالمبحوثين لاخذ اجاباتهم، وتعد الطريقة ملائمة للحالات التالية :

(1) اذا كان عدد المشمولين (عدد الاستبانات) صغيرا .

(2) اذا كان بعض او نسبة مهمة من المبحوثين اميين .

(3) اذا كانت اسئلة الاستبيان تحتاج الى شرح وتوضيح يصعب شرحها في التعليمات.

ويلعب الباحث دورا مهما في دقة المعطيات التي تجمع بهذه الطريقة من خلال اسلوب تعامله اثناء مقابلة المبحوثين، مما يستدي العناية باختيار الباحثين الذين يقومون بمهمة ملئ الاستبانات.

ميزات وعيوب طريقة المقابلة الشخصية :

من ميزات الطريقة انها تساعد المبحوثين من خلال الشرح والتوضيح الذي يقدمه الباحث خلال المقابلة، مما يساعد على زيادة دقة المعطيات وتقليل نسبة الخطأ فيها.

اما عيوبها فتتمثل بحاجتها لاعداد كبيرة من الباحثين وزيادة التكاليف، كما انها قد تؤدي الى تحيز الباحث من خلال تاثيره الشخصي .

2- 6- 4 طريقة الهاتف Telephone Method

وهي قليلة الاهمية والاستخدام، ويفضل استخدامها في الحالة التي تكون فيها المعطيات المستهدفة محدودة وتتعلق باستطلاع اراء عن شان اجتماعي او اقتصادي محدد وذات فائدة مباشرة للمبحوثين، كأن تتعلق بتخفيض الضرائب مثلا .

2- 6- 5 طريقة التركيز على المناقشات الجماعية
Focus Group Discussion Method

وهي حديثة الاستخدام وتتسم بالشفافية، وفحواها اثارة الاهتمام بصورة غير مباشرة عن موضوع معين في الاماكن العامة كالنوادي او اماكن العمل او المقاهي وما شابه للاطلاع على اراء الجماعة المعنية بالامر بصورة عفوية مجردة من التاثيرات. والطريقة تكون مناسبة في الحالات التي يصعب فيها ادلاء المستجوب بالمعلومة الصحيحة امام المعنيين او الموظفين الرسميين خشية من رب العمل او لاحساسة بتحفظ من الافصاح عن رايه حفاظا على عمله او تحاشيا من عواقب قد تكون موجودة او غير موجودة. كما في حالة الاستفسار مثلا من عاملين في مصنع ما عن ظروف العمل الصحية والغذائية او عن جودة الخدمات المقدمة لهم في العمل وكانت هكذا خدمات وظروف عمل ليس بالمواصفات المطابقة بموجب التعليمات والقوانين .

الفصل الثالث

تصميم العينات واسلوب تحديد حجمها

Sampling Design and Sample Size Determination

3-1 مفهوم تصميم العينة

عندما يتقرر أجراء الاستقصاء او المسح الإحصائي بأسلوب العينة، فأن ذلك يعني ان توفير المعطيات عن خصائص المجتمع سيعتمد على جزء من هذا المجتمع، ويطلق على هذا الجزء بالعينة، كما ويشترط في العينة أن تكون ممثلة لخواص المجتمع كافة بالنسبة للظاهرة التي نقوم بدراستها بما في ذلك الاختلاف بين وحداته وبحدود ما يسمح به حجم العينة تبعا لمقياس الدقة والإمكانيات المتاحة للدراسة. ان تحديد هذا الجزء وعملية اختيار وحداته يتم إنجازها من خلال ثلاثة مراحل هي:

المرحلة الاولى: تحديد نوع العينة Type of Sample، بالاعتماد على طبيعة المجتمع الإحصائي وخصائصه من ناحية درجة تجانس وحداته من عدمها، وعلى مدى توفر الإطار الإحصائي للمجتمع.

المرحلة الثانية: تحديد حجم العينة Sample Size، ويراعى في اختيار أداة تحديد حجم العينة، طبيعة الخاصية تحت الدراسة إن كانت على شكل نسبة أو قيمة مطلقة، وفيما إذا كان تباين المجتمع متوفر أم لا.

المرحلة الثالثة : تحديد طريقة اختيار وحدات العينة Selection Method of Sample Units، ومن بين أساليب عملية الاختيار هو الأسلوب

الدوري Periodic Method من خلال توظيف العينة العشوائية النظامية خاصة في حالات العينات الطبقية والعنقودية، وطريقة السحب العشوائي Random Selection Method. والطريقتين متوفرة في كل من برنامجي SPSS و EXCEL .

والعينات على نوعين هما: العينات العشوائية (الاحتمالية) والعينات غير العشوائية (غير الاحتمالية) .

2-3 مفهوم العينات العشوائية

Random Samples Definition

وهي العينات التي يجب ان تكون مستوفية للشروط التالية :

الشرط الاول : كل عينة يتم اختيارها من المجتمع لها احتمال معلوم، وتبعا لذلك كل وحدة في العينة يجب ان يكون لها ايضا احتمال معلوم لكي يتم شمولها في العينة. وليس من الضروري ان يعني هذا الاحتمال المعلوم تساوي الاحتمال لكل وحده في المجتمع كما هو الحال في العينات العشوائية البسيطة Simple random sample، بل قد يختلف وهذا الاختلاف يساعد في حالة المجتمعات غير المتجانسة على توفير دقة أعلى للتقديرات التي نحصل عليها من العينة كما سيتضح عند التطرق فيما بعد إلى العينات العشوائية الطبقية Stratified random sample .

الشرط الثاني: ان يتم سحب العينة باستخدام إحدى طرق الاختيار العشوائي، بحيث تتحقق الاحتمالات المعلومة الواردة في الشرط الاول أعلاه.

الشرط الثالث: ان يتم اعتماد الاحتمالات المعلومة عند استخدام نتائج العينات في الحصول على تقديرات جيدة لمعالم المجتمع الذي نقوم بدراسته.

والعينات الاحتمالية او العشوائية على عدة أنواع، يعتمد ويتوقف استخدام كل نوع منها على طبيعة المجتمع والغرض من الدراسة والإمكانات المتاحة للبحث، وستعرض فيما يلي بإيجاز لأهم هذه الأنواع وطرق استخدامها.

3-3 العينة العشوائية البسيطة Simple Random Sample

3- 3- 1 مفهوم العينة وشروطها

وهي العينة التي يتم اختيارها بطريقة تعطي لكل وحده واحدة من المجتمع الإحصائي N فرصة الظهور نفسها في كل مرة من مرات الاختيار (1/N)، وبذلك فلكل عينة حجمها n احتمال الاختيار نفسه من بين العينات الممكنة أي :

$$\frac{1}{\binom{N}{n}}$$

إذ إن الصيغة أعلاه تمثل عدد العينات الممكن اختيارها بحجم n من مجتمع حجمه N ونحصل عليها باستخدام صيغة التوافيق combination الآتية :

$$\binom{N}{n} = \frac{N!}{n!\ (N-n)!}$$

حيث إن : !N تدعى عاملي N (مضروب N) ومفكوكة هو :

$$(N)\ (N\text{-}1)\ (N\text{-}2) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (2)\ (1)$$

مثال (3-1): إذا كان لدينا مجتمع إحصائي متكون من الوحدات الآتية: B, C, D, E فان عدد العينات الممكن سحبها لحجم n = 2 باستخدام الصيغة اعلاه تتكون من 6 عينات هي : BC, BD, BE, CD, CE, . DE

ونلاحظ إن لكل من هذه العينات لها نفس الاحتمال وهو 1/6 وان لكل وحده في المجتمع لها الاحتمال نفسه في الظهور وهو 3/6 = 1/2. من ذلك نستدل على ان العينة العشوائية البسيطة لها صفتان أساسيتان هما: إن لكل عنصر (أو وحده) في المجتمع احتمال الظهور نفسه، وان لكل من العينات الست لها أيضا احتمال الاختيار نفسه .

3- 3- 2 حالات استخدام العينة العشوائية البسيطة

تستخدم العينة العشوائية البسيطة عندما يكون المجتمع متجانسا من حيث الغرض أو الصفة التي تتعلق بها الدراسة، وهي من أبسط أنواع العينات و تعد أساسا لاختيار كل هذه الانواع .

3- 3- 3 أساليب اختيار العينة Sample Selection Methods

(1) الاختيار بالإرجاع (Selection With Replacement) وهو يعنى أننا حين نختار مفرده من المجتمع فأننا نعيدها ثانيه إلى المجتمع ليتم اختيار المفردة الثانية، وقد تظهر المفردة نفسها أو غيرها .

(2) الاختيار بدون إرجاع (Selection Without Replacement) وهو يعنى انه عند اختيارنا للمفرده الأولى فأننا لا نلجأ إلى إعادتها ثانيه إلى المجتمع وأنما نختار مفرده مما تبقى من المجتمع وهكذا. ومن الناحية العملية فان جميع مسوحات العينة تعتمد على اسلوب الاختيار بدون إرجاع، لذا سيكون التركيز على هذا الاسلوب في دراستنا للعينات.

3- 3- 4 اساليب الاختيار العشوائي لوحدات العينة
Random Selection Method Of Observations

كما هو الحال مع جميع المجالات، فقد شملت عملية التوسع في استخدام الحاسب الآلي، إجراءات السحب العشوائي لوحدات العينة، واصبح

بالإمكان في حالة إدخال معطيات المجتمع إلى الحاسوب في الحصول على العينة من خلال استخدام الايعازات التالية :

برنامج SPSS ⇦ تحليل Analysis ⇦ العينة Sample ⇦ طريقة السحب (periodic or random)

او برنامج Excel ⇦ ادوات Tools ⇦ تحليل البيانات (Data analysis) ⇦ المعاينة (Sampling) ⇦ طريقة السحب (periodic or random)

حيث تتحقق بهذا الإيعاز عملية سحب العينة وهي أما الدورية periodic باعتماد أسلوب العينة العشوائية المنتظمة والتي تعتمد العشوائية في جزئها الأول، أو طريقة السحب العشوائي المباشر Random. إلا انه في بعض الاحيان وخاصة عندما يكون حجم المجتمع صغير او محدود قد يتم اللجوء إلى الطريقة اليدوية التقليدية في استخدام جداول الأرقام العشوائية Random Numbers Tables والمبين نموذج منه في الملحق رقم (1.3)، والتي تحتوى على أرقام تم الحصول عليها بطريقه عشوائية، اى بطريقه غير خاضعة لأي نوع من أنواع الترتيب والتي تتلخص بالخطوات التالية :

(1) نعطى أرقاما متسلسلة لعناصر (وحدات) المجتمع المراد دراسته.

(2) تحديد عدد الأعمدة التي سنستخدمها من الجدول العشوائي للحصول على الأرقام المطلوبة، ويتوقف هذا على حجم المجتمع. فبذلك نختار عدد الأعمدة بحيث يكون مساويا لعدد خانات اكبر رقم أعطي للمجتمع .

(3) نحدد نقطه البداية في الجداول العشوائية .

(4) نبدأ باختيار أول رقم من الجدول من نقطه البداية التي حددناها شرط ان يكون من ضمن الأعمدة التي اخترناها، فالعدد الذي يليه في هذه

الأعمدة يكون الرقم الثاني وهكذا إلى ان نحصل على عدد وحدات العينة المطلوبة مع استبعاد أي عدد يتكرر او أي عدد اكبر من عدد عناصر (مراتب Digits) المجتمع الإحصائي، اي اذا كان حجم المجتمع اقل من 100 نعمل على مرتبتين واكثر من 100 الى اقل من 1000 نعمل على ثلاثة مراتب وهكذا.

(5) نحدد عناصر المجتمع التي تحمل الأرقام المختارة لتكون وحدات العينة العشوائية البسيطة المراد اختيارها من هذا المجتمع .

مثال (2.3) : إذا كنا بصدد القيام بدراسة عن أوضاع العاملين في أحد المصانع وكان مجموعهم 500 عامل والمطلوب اختيار عينه عشوائية حجمها 10%، حدد وحدات العينة باستخدام جداول الأرقام العشوائية .

الحل ل (2.3) :

- بما أن عدد العاملين هو 500 وان حجم العينة المطلوبة يمثل نسبة قدرها 10% فان حجمها هو $n = 50$ عاملا، وبذلك نعطى أرقاما لجميع العاملين من 1 إلى 500

بما ان اكبر عدد أعطي لوحدات المجتمع هو 500 يتكون من ثلاثة مراتب (خانات) أذن يكون عدد الاعمده التي سنستخدمها كل مره هو 3 أعمده (أي ان كل عدد يتكون في ثلاثة أرقام).

- نحدد نقطه البداية في جدول الأرقام العشوائية، ولتكن بداية الجدول في الملحق(1.3) ولثلاث مراتب فنجد أنه الرقم 809 ولما كان هذا الرقم اكبر من 500 عليه يتم إهماله ونأخذ الرقم الثاني وهو 356 وبما انه اقل من 500 فأن علينا عده الرقم الأول في العينة. ثم نأخذ الرقم الثاني المكون أيضا من ثلاث مراتب وهو 133 وبما أنه أقل من حجم المجتمع 500 فهو يعد الرقم الثاني في العينة وهكذا حتى نحصل على 50 رقما

من بين لـ500 دون تكرار لأي منها، وبموجب ذلك فأن أرقام العينة هي:

297, 313, 39, 65, 470, 400, 449, 358, 133, 356, 458, 340, 465, 280, 408, 405, 232, 63, 82, 297, 425, 276, 480, 350, 496, 216, 298, 233, 443, 104, 258, 382, 468, 228, 423, 397, 410, 319, 332, 287, 328, 110, 439, 487, 323, 141, 135, 191, 161, 121.

- الآن نحدد أسماء العاملين الذين يحملون هذه الأرقام ليكونوا هم وحدات العينة العشوائية البسيطة المطلوبة .
- يمكن الحصول على المعطيات المطلوبة للدراسة من وحدات هذه العينه .
- تعمم النتائج التي نحصل عليها من هذه العينه على مجتمع العاملين بالمصنع كله وذلك باعتبار أن المعطيات التي حصلنا عليه من العينه تعد ممثله لجميع العاملين في المصنع .

مثال (3.3): لدينا مجتمع إحصائي مكون من 50 مخزن لبيع المواد الغذائية، وكانت قيم المبيعات اليومية (بلدولار) لهذه المخازن هي :

112, 132, 132, 131, 080, 126, 116, 118, 073, 130
116, 128, 062, 132, 091, 127, 118, 132, 132, 084, 124,
190, 109, 112, 090, 117, 127, 234, 119, 121, 128, 087,
087, 132, 129, 119, 122, 114, 093 ,123 , 131,126, 112
,089, 121, 118, 116, 136, 119,120

والمطلوب اختيار 10 وحدات (مخازن) كعينه عشوائية بسيطة .

الحل لـ (3.3) :

- على وفق الخطوات الواردة في الفقرة (4) من البند (3.3) نقوم بترقيم وحدات المجتمع الإحصائي من 1 إلى 50 والتي تتكون من مرتبتين.

- نستخدم الجدول في الملحق رقم (1.3) مبتدئين من السطر الأول عند العمود الثاني لتحديد وحدات العينة التي يتم سحبها. فتظهر لنا الأرقام الآتية : 48 , 35 , 49 , 21 , 29 , 23 , 44 , 10 , 45 , 03
- وحسب تسلسل قيم المبيعات الواردة في المثال، نجد أن هذه الأرقام تعود إلى القيم الآتية :

132، 089، 130، 112، 190، 234، 084، 136، 132 116

وهى تمثل وحدات العينة العشوائية البسيطة .

3- 3- 5 عيوب العينة العشوائية البسيطة وميزاتها

تظهر عيوب العينة العشوائية البسيطة في المجالات الآتية : -

(1) إذا كانت وحدات المجتمع غير متجانسة في الصفة التي نقوم بدراستها، فأن استخدام العينة العشوائية لا يضمن ان تكون العينة ممثله لهذه الصفة بالمجتمع .

(2) في حالة كون المجتمع الإحصائي كبيرا، فأن استخراج وحدات العينة العشوائية يحتاج إلى مجهود كبير لتهيئه إطار المجتمع وبخاصة إذ لم نستخدم في العملية الحاسب الآلي .

(3) عندما تكون وحدات العينه موزعه على مناطق جغرافية واسعة ومتباعدة فأن تكاليف جمع المعطيات من هذه الوحدات تكون عالية عادة مع صعوبة أحكام الإشراف على العمل الميداني. وفى الواقع غالبا ما تعالج هذه العيوب باستخدام إحدى العينات العشوائية الأخرى التي سنشرحها لاحقا .

اما ميزات العينة :- فتتمثل بكون العينة العشوائية البسيطة وكما ذكرنا تعد الأساس لباقي أنواع العينات فضلا عن كونها من أبسط هذه العينات استخداما .

4-3 تقدير معالم المجتمع باستخدام نتائج العينة العشوائية البسيطة

بعد أن توصلنا إلى معرفة كيفية الحصول على العينة العشوائية البسيطة، ياتي السؤال الآن عن كيفية تعميم نتائج العينة على المجتمع الإحصائي الذي سحبت منه، لاسيما وان نتائجنا مبنية على عينة فقط. فمثلا عند استطلاعنا لاراء عينة من الأشخاص تمثل 100 مسافر عن مستوى خدمات النقل العام، أو عينه من الطلبة تتكون من 80 طالبا عن الهوايات التي يمارسونها في أوقات فراغهم فان مدار اهتمامنا الحقيقي لن يكون البحث عن آراء هؤلاء لـ 100 شخص عن مستوى خدمات النقل ولا أراء هؤلاء لـ 80 طالبا عن الهوايات التي يمارسوها. ولكن لتقدير أراء المجتمع الإحصائي لكافه المسافرين في الحالة الأولى ولجميع الطلبة في الحالة الثانية.

أذن فأن المسألة الأساسية للعينات هي إيجاد تقديرات تكون ممثله لمعالم المجتمع Population parameters عموما مثل المتوسط ونسبه خاصية معينه بالمجتمع والمجموع الكلى وتباين كل منها، وأن مثل هذه التقديرات تكون ممكنة مع العينات العشوائية. وذلك لانه بالإمكان استخدام نظريه الاحتمالات التي تقوم على أساسها هذه العينات في دراسة الأخطاء المعيارية للتقديرات وللتحيز ولإيجاد حدود فترة الثقة لقيم معالم المجتمع الحقيقية بالاستناد إلى نتائج العينة. وبذلك نتمكن من تكوين فكره عن مدى دقه ومقبولية هذه التقديرات.

3- 4- 1 انواع القيم التقديرية لمعالم المجتمع للعينة العشوائية البسيطة.

من الناحية التطبيقية، فإن الاستدلال بالقيم التقديرية لمعالم المجتمع من نتائج العينة، يمكن ان يأخذ نوعين من التقديرات هي :

(1) التقدير بنقطة **Point Estimation**

ويتمثل التقدير هنا في نقطه أو قيمه محدده كقيمة الوسط الحسابي أو قيمة الانحراف المعياري وعدها تقديرا لمعالم المجتمع، أن تباين العينة S^2 و انحرافها المعياري s تجهزنا بالأساس لتقدير تباين المجتمع σ^2، إذ إن معلمة التباين من نتائج العينة هو عبارة عن تباين متوسط العينة وهو :

$$\sigma\frac{2}{x}=\sigma^2/n$$

والخطأ المعياري يقدر بالاحصاءه :

$$S_{\bar{x}}=S/\sqrt{n}$$

فإذا كانت لدينا عينه حجمها $n = 6$ وقيم وحداتها هي: 1، 2، 4، 5، 7، 11 مسحوبة من مجتمع مجهول المعالم. فإن الوسط الحسابي للعينة $\bar{x}$ هو المقدر Estimator وان قيمته التي هي 5 هي نقطة التقدير، أي تقدير لمتوسط المجتمع μ غير المعلوم .

وإذا أفترضها أن نسبة عدد القيم الزوجية للعينة وهى p (حرف حجم صغير) هي المقدر لنسبه المجتمع P (حرف حجم كبير)، فان $\rho=1/3$ هي نقطه التقدير، أي قيمة المقدر لنسبة المجتمع P. وكذا يمكن القول عند أستخدام S كتقدير للانحراف المعياري للمجتمع غير المعلوم σ .

(2) تقدير الفترة **Interval Estimation**

وفيها يمكننا تقدير المقدر بفترة يصاحبها مقدار ثقة معلوم إذ يقصد بالثقة هنا مقدار الاحتمال الذي نثق (نضمن) به إن الفترة (المدى) تشمل قيمة المعلمة المجهولة μ. وقد سمي هذا الاحتمال بمعامل ثقة. فعندما نقول مثلا بثقة مقدارها 95% فذلك يعنى ان هناك فرصه مقدارها 95 من 100 بأن الفترة تتضمن قيمة المتوسط الحقيقي للمجتمع μ . وتشتمل فتره الثقة

على حدين هما: الحد الأدنى lower limit والحد الأعلى upper limit لفترة الثقة، إذ نتوقع أن تكون قيمه المتوسط للمجتمع ضمنها وذلك باحتمال معلوم. وبما أن العينة تتكون من جزء صغير من المجتمع فأن من الصعب التأكد 100% من صحة هذه الفترة وعليه فان حساب مدى فتره الثقة سوف يعتمد على مقدار معامل الثقة. فمعامل ثقة 95% مثلا يعنى أننا نتوقع في 95% من الحالات أن معلمه المجتمع ستكون بين هذين الحدين وان هناك 5% من الحالات تكون خارجها. كأن نقول مثلا أن اختيار الصيغة المناسبة لبناء حدي فتره ثقة بصوره عامه يعتمد على التوزيع الذي تتبعه الاحصاءه Statistic وسوف نتعرض هنا لكيفية حساب فترة الثقة للمتوسط والتباين .

2-4-3 فترة الثقة للمتوسط، Confidence Interval For μ

(1) فتره الثقة لمتوسط مجتمع طبيعي تباينه معلوم .

عند اختيار عينة عشوائية من مجتمع طبيعي انحرافه المعياري σ معلوم، نتوقع بأن يكون متوسط العينة $\bar{x}$ هو أيضا يتبع التوزيع الطبيعي normal distribution بخطأ معياري $\sigma/\sqrt{n}$ إذ أن n هو حجم العينه. وعليه فعند ثقة 95% فأن المنطقة التي يقع ضمنها المتوسط الحقيقي μ تقل بمقدار $1.96\ \sigma/\sqrt{n}$ عن المتوسط الحقيقي وتزداد بمقدار $1.96\ \sigma/\sqrt{n}$، وكما هو موضح في الشكل البياني رقم (1.3)، بعبارة أخرى فأن المتغير الطبيعي المعياري Standard normal variable لتوزيع متوسط العينة Z سيكون:

$$Z = \bar{x} \mp \mu/\sigma_{\bar{x}}$$

حيث أن :

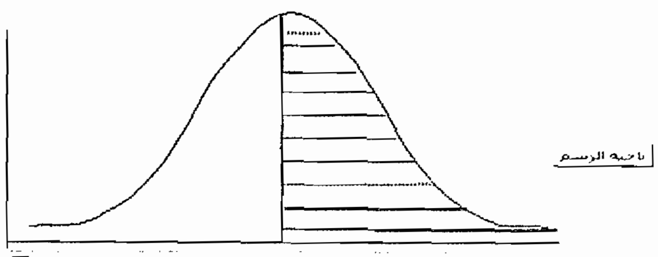
$\bar{x}$ = متوسط العينة

μ = متوسط المجتمع

σ_x = الانحراف المعياري للمتوسط، أي $\sigma/\sqrt{n}$

شكل بياني رقم (1.3)

يمثل القيمة الاحتمالية للقيم الطبيعية المعيارية الواقعة بين 0 و 1.96



إذ أن 95% من متوسط العينة $\bar{x} \pm 1.96\ \sigma_{\bar{x}}$ عن متوسط المجتمع μ وهكذا فان متوسط المجتمع سيكون أيضا ضمن $\bar{x} \pm 1.96\ \sigma_{\bar{x}}$ من متوسطات العينة عند ثقة مقدارها 95%، أي المتوسط الحقيقي μ سيكون ضمن الفترة $\bar{x} \pm \sigma/\sqrt{n}$ بثقة مقدارها 95% وبصورة عامه إذا كانت درجة الثقة (معامل الثقة Confidence coefficient) هي $(1-\alpha)100\%$ فإن بالإمكان الحصول على فترة الثقة من الصيغة الآتية :

$$\bar{x} - Z_{(1-\alpha/2)}\sigma/\sqrt{n} \leq \mu \leq \bar{x} + Z_{(1-\alpha/2)}\sigma/\sqrt{n}$$

إذ يدعى $\bar{x} - z_{(1-\alpha/2)}\sigma/\sqrt{n}$ بالحد الأدنى لفترة الثقة

و $\bar{x} + z_{(1-\alpha/2)}\sigma/\sqrt{n}$ بالحد الأعلى لفترة الثقة،

و z هي القيمة المعيارية المناظرة لدرجه الثقة 100% (1-α) . اي α هي الاحتمال المكمل لدرجه الثقة، وعادة ما يطلق عليها مستوى المعنوية في موضوع اختبار الفروض.

مثال (4.3): مجتمع موزع توزيعا طبيعياً، انحرافه المعياري $\sigma = 11$ سحبت منه عشوائيا عينة حجمها $n = 26$ وكان متوسط العينة $\bar{x} = 48$ أحسب تقدير متوسط المجتمع μ بثقة مقدارها 95% .

الحل لـ (4.3) :

- نستخرج قيمتا كل من الانحراف المعياري للمتوسط و القيمة الجدولية لـ $Z_{(1-a/2)}$ وكالآتي :

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}$$
$$= 11/\sqrt{n} = 2.157$$

ومن الملحق رقم (4.3) نجد إن : $Z_{(1-\alpha/2)} = Z_{.975} = 1.96$:

- فنحصل على :

$$\bar{x} - z_{(1-\alpha/2)}\sigma/\sqrt{n} \le \bar{x} + z_{(1-\alpha/2)}\sigma/\sqrt{n}$$
$$48 - (1.96)(2.157) \le \mu \le 48 + (1.96)(2.157)$$
$$43.77 \le \mu \le 52.23$$

أي عند ثقة مقدارها 95% ستكون قيمة متوسط المجتمع μ بين القيمتين (52.23 : 43.77)

(2) فتره الثقة لمتوسط المجتمع μ غير معلوم التباين σ^2.

إذا كان تباين المجتمع الطبيعي غير معلوم حينئذ يتم تقديره من نتائج العينة، وذلك من خلال احتساب الخطأ المعياري التقديري لمتوسط العينة.

$$s_{\bar{x}} = s/\sqrt{n}$$

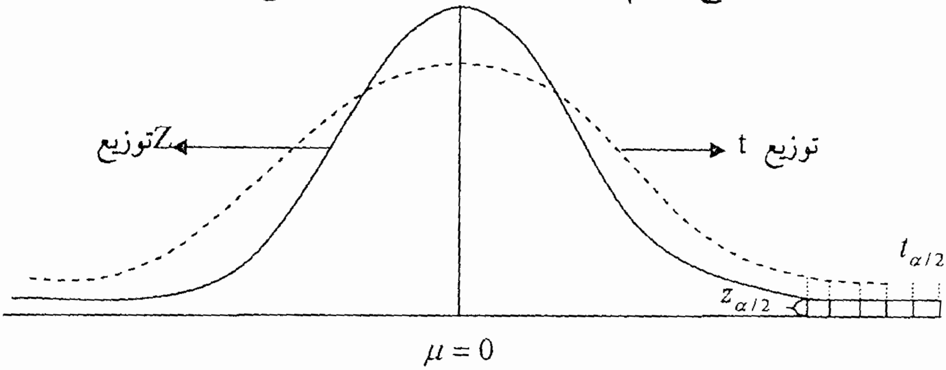
وفى مثل هذه الحالة التي نستبدل فيها $\sigma_{\bar{x}}$ بـ $s_{\bar{x}}$. ونستخدم قيمة توزيع t بدلا من z على النحو الآتي :

$$\bar{x} - \mu/(s_{\bar{x}}) = \bar{x} - \mu/(s/\sqrt{n})$$

وكما هو الحال مع z فإن توزيع t سيكون متماثلا أيضا مع متوسطه الذي يساوي الصفر ويعتمد على حجم العينة n، لكن أطرافه تكون مرتفعه اكثر مما عليه في حالة توزيع z، لذلك فان قيمة t تكون أكبر من قيمه z للاحتمال نفسه، وكما مبين في الشكل البياني رقم (2.3) في ادناه :

شكل بياني رقم (2.3)

يوضح حجم كل من z و t عند طرفي التوزيع الطبيعي



لذا فان مساحة توزيع t تكون مجدولة حسب درجات الحرية v حيث ان v = n - 1 وكلما ازداد عدد درجات الحرية اقتربت قيمة t الجدولية من قيمة z الجدولية. ونتيجة لذلك تستبدل ($z_{(1-\alpha/2)}$) بـ ($t_{(1-\alpha/2)}$، v) لتصبح صيغة فتره الثقة في حالة مجهولية تباين المجتمع كما يأتي :

$$\bar{x} - t_{(1-\alpha/2),v}\ s/\sqrt{n} \le \mu \le \bar{x} + t_{(1-\alpha/2),v}\ s/\sqrt{n}$$

مثال (5.3): قدر قيمة متوسط المجتمع μ الموزع طبيعيا، المسحوبة منه عينه حجمها n = 10 وقيم وحداتها كالآتي: 12،5،13،8،6،11،15،10،7،8. وعند درجة ثقة مقدارها 95%.

الحل لـ (5.3) :

- نحسب قيمه متوسط العينة $\bar{x}$، ونقدر قيمة الانحراف المعياري للمجتمع σ بواسطة الانحراف المعياري للعينة s كالآتي :

$$\bar{x} = 9.5$$
$$s = 3.24$$

- نحدد درجات الحرية v والقيمة الجدولية لـ t المناظرة لدرجة ثقة 95% من الملحق رقم (5.3) وهكذا نجد أن :

$$v = n - 1 = 9$$
$$t_{(1-\alpha/2)},_v = t_{0.975},_9 = 2.26$$

- فنحصل على :

$$\bar{x} - t_{(1-\alpha/2)},_v\ s/\sqrt{n} \le \mu \le \bar{x} + t_{(1-\alpha/2)},_v\ s/\sqrt{n}$$

$$9.5 - (2.26)\ 3.24/\sqrt{10} \le \mu \le 9.5 + (2.26)\ 3.24/\sqrt{10}$$
$$7.2 \le \mu \le 11.8$$

من ذلك نستدل على إن متوسط المجتمع يقع بين القيمتين 7.2 و 11.8 بثقة مقدارها 95% .

(3) فترة الثقة لنسبة خاصية المجتمع P

أما إذا كانت قيم العينة من توزيع ذات الحدين Binomial distribution فأن الوسـط الحسـابي $p=\sum xi / n$ والوسط الحسـابي للتـوزيع $P = \mu$، فان p تخضـع للتوزيع الطبيعي التقريبي عنـدما $n > 30$، أي أن $b(1, p)$ يصبح $N(P, PQ/n)$.

حيث ان :

$b(1, p)$ تشير إلى التوزيع ذات الحدين

$N(P, PQ/n)$ تشير إلى التوزيع الطبيعي التقريبي.

وبذلك فان فترة الثقة $(1-\alpha)100\%$ للنسبة P هي :

$$P - Z_{(1-\alpha/2)}, / \sqrt{pq\ /n} \leq\ p\ P + Z_{(1-\alpha/2)} / \sqrt{pq\ /n}$$

حيث إن $q = 1 - p$

مثال (س 6.3): في مسح إحصائي لمدينة ما أخذت عينة حجمها = n 300 شخص، وجد بينهم 123 من الذكور. والمطلوب ايجاد فترة الثقة عند 95% لنسبة الذكور في المجتمع P .

الحل لـ (6.3) :

- نحسب قيمة نسبة العينة p وهي : $p = 123/300 = 0.41$

ومن الملحق رقم (4.3) نجد ا ن : $Z_{0.975} = 1.96$

- فنحصل على :

$$p - Z_{(1-\alpha/2)}, / \sqrt{pq\ /n} \leq\ P\ \leq p + Z_{(1-\alpha/2)} / \sqrt{pq\ /n}$$

$$0.41-(1.96)(0.028) \leq P \leq 0.41+(1.96)(0.028)$$

$$0.355 \leq P \leq 0.465$$

أي أن فترة الثقة لنسبة الذكور في المجتمع P تقع بين القيمتين 0.355 و 0.465 عند درجة ثقة مقدارها 95% .

3-4-3 فتره الثقة لتباين المجتمع σ^2 .

بالإضافة إلى حصولنا على فترة ثقة لمتوسط المجتمع ونسبة خاصية معينة، فإن من المفيد أن نحصل من نتائج العينة على فتره ثقة لتباين المقدرات Estimators التي تم الحصول عليها، ولوضع صيغه فتره ثقة لهذا الغرض نستخدم توزيع مربع كاي χ^2 وهو من التوزيعات المتصلة، حيث تأخذ χ^2 القيم من صفر إلى ما لانهاية في الاتجاه الموجب. ويعتمد شكل التوزيع على

حجم العينة n وبدرجات الحرية v = n - 1، ويصبح شكل التوزيع مشابها للتوزيع الطبيعي كلما زاد حجم العينة n ، ففي حالة عينة عشوائية قيم وحداتها (x_i)

حيث ان ، i =1، 2، 3،....، n مسحوبة من مجتمع طبيعي متوسطة μ وانحرافه المعياري σ، فلكل عينة من مجموع العينات التي يمكن تشكيلها بحجم n نستخدم صيغة المتغير العشوائيس الطبيعي المعياري z ليصبح مجموع مربعات المتغير على هذا النحو :

$$\sum Z^2 = \sum \frac{(X-\mu)^2}{\sigma^2}$$

$$= \frac{1}{\sigma^2} \sum (x_i - \mu)^2$$

وعند تقدير متوسط المجتمع μ بمتوسط العينة $\overline{X}$ يصبح لدينا :

$$\sum Z^2 = \frac{1}{\sigma^2} \sum (x_i - \overline{x})^2$$

وهو عبارة عن مجموع مربعات انحرافات المتغير عن وسطها الحسابي مقسوما على معامل ثابت وهو التباين σ^2.

حيث ان :

$$\sum (x_i - \overline{x})^2 = (n-1)s^2$$

وبالامكان التعبير عن المعادلة اعلاه بالصيغة التالية :

$$\frac{(n-1)s^2}{\sigma^2}$$

وهو يتبع في تغيراته توزيع مربع كاي χ^2 بدرجات حرية v=n-1

وباستخدام توزيع χ^2 فان صيغة فترة الثقة للتباين σ^2 تكون :

$$P\left[\frac{(n-1)s^2}{\chi^2_{(1-\alpha/2),v}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{\alpha/2,v}}\right] = 1-\alpha$$

حيث إن :

$\chi^2_{(\alpha/2)}$، v هي قيمة مربع كاي الجدولية بدرجات حرية مقدارها v التي تحصر احتمالا قبلها مساويا $\alpha/2$.

$\chi^2_{(1-\alpha/2)}$،v التي تحصر احتمالا قبلها قدره $(1-\alpha/2)$. ويتم الحصول على هذه القيم من الملحق رقم (6.3) .

وأن :

$\frac{(n-1)s^2}{\chi^2_{(1-\alpha/2),v}}$ هو الحد الأدنى لفترة الثقة .

$\frac{(n-1)s^2}{\chi^2_{\alpha/2,v}}$ هو الحد الأعلى لفترة الثقة .

وباستخدام المعادلة $\Sigma(x_i-\bar{x})^2 = (n-1)s^2$ يمكن كتابه فتره الثقة على النحو الآتي :

$$\frac{\Sigma(x-\bar{x})^2}{\chi^2_{(1-\alpha/2),v}} \le \sigma^2 \le \frac{\Sigma(x-\bar{x})^2}{\chi^2_{\alpha/2,v}}$$

مثال (7.3) : نفترض أن لدينا عينه عشوائية بحجم $n = 7$ وأن قيم وحداتها هي: 2 ،3 ،7 ،5 ،9 ،6 ،4 مسحوبة من مجتمع طبيعي. فما فترة الثقة التي تتضمن تباين المجتمع σ^2 بدرجه ثقة 95%.

الحل لـ (7.3) :

- نحسب مجموع مربعات الفروق $\sum(x_i-\bar{x})^2$ أي :

$$\bar{x} = \sum x_i / n$$
$$= 5.143$$

$$\sum(x_i-\bar{x})^2=(2-5.143)^2+\ldots+(4-5.143)^2$$
$$=34.857$$

- نجد قيمه كل من: $\chi^2_{(1-\alpha/2}, \chi^2_{(\alpha/2))}$ عند درجات حرية n-1=6 وباستخدام الجدول في الملحق رقم (6.3) وعند درجة عدم ثقة مقدارها 0.025 و 0.975 نجد أن :

$$\chi^2 0.025,6 = 14.449$$

$$\chi^2 0.975,6 = 1.237$$

- وبتطبيق صيغة فترة الثقة نحصل على :

$$\frac{\Sigma(x-\bar{x})^2}{\chi^2_{(1-\alpha/2),V}} \le \sigma^2 \le \frac{\Sigma(x-\bar{x})^2}{\chi^2_{\alpha/2,V}}$$
$$\frac{34.857}{14.412} \le \sigma^2 \le \frac{34.857}{1.237}$$
$$2.412 \le \sigma^2 \le 28.178$$

وهكذا فأن الفترة تشمل تباين المجتمع σ^2 بين القيمتين 2.412 و28.178عند درجة ثقة مقدارها 95% .

4-4-3 حالة رقم دراسية C_{3-1} :

في توزيع المعاينة Sampling Distribution

(إثبات مدى تكافؤ إحصاءات العينة مع معالم المجتمع المسحوبة منه العينة)

من المفيد أن نشير في البدء إلى إن هناك ثلاثة أنواع من التوزيعات هي :

(1) توزيع المجتمع **(Population distribution)** وهى الحالة التي يمكننا فيها معرفة جميع وحدات المجتمع وبما يتيح لنا وضعها في جدول توزيع تكراري ونحصل منه مباشرة على الوسط الحقيقي المجتمع µ أو على أي معلمة أخرى.

(2) توزيع العينة (Sample distribution) والمتمثل في اختيار عينه من المجتمع ووضع معطيات هذه العينة في توزيع تكراري، فالوسط الحسابي $\bar{x}$ لهذه العينة لا يكون مساويا للوسط الحسابي الحقيقي للمجتمع μ وكذا الحال بالنسبة للمقاييس الأخرى كالتباين مثلا .

(3) توزيع المعاينة (Sampling distribution) ويعنى الحصول على جميع العينات الممكنة من المجتمع والتي تكون من نفس الحجم n فإذا ما وجدنا الوسط الحسابي لكل من هذه العينات ووضعنا هذه المتوسطات في توزيع تكراري فأن هذا التوزيع يسمى بتوزيع المعاينة للمتوسطات الحسابية. لكن سيكون من المستحيل إيجاد جميع العينات عندما يكون حجم العينة كبيرا جدا أو لانهائيا، لذلك يتم اشتقاقه رياضيا. إن إمكانية تقدير أدلة المجتمعات عن طريق دراسة هذه العينات وتحليلها هو ما يسمى بالاستدلال الإحصائي (اوالاستنباط الاحصائي) Statistical inference أي استخدام نظرية الاحتمال في إصدار قرارات عن مجتمع أو عدد من المجتمعات عن طريق عينات مأخوذة منها بشرط أن تكون العينات عشوائية مسحوبة من مجتمع توزيعه طبيعي. ولتوضيح كيفية الحصول على توزيع المعاينة Sampling Distribution من مجتمع صغير نسبيا ومدى تكافؤ المقاييس بين كل من المجتمع والعينة في الحالة التي تكون فيها مسحوبة من مجتمع طبيعي محدود $N(\mu,\sigma^2)$ نتابع المثال الآتي :

مثال (8.3) : بفرض لدينا مجتمعا احصائيا يتكون من 5 وحدات هي: a، b،c، d،e، وقيم هذه الوحدات هي: 18،3،6،3،0 على التوالي، والمطلوب اثبات تكافؤ كل من احصاءات العينة ومعالم المجتمع عند سحب عينة n=3

▪ لدينا متوسط وتباين المجتمع وهي :

$$\mu = \sum X_i / N$$
$$= 30 / 5 = 6$$
$$\sigma^2 = \sum (X_i - \mu)^2 / N$$
$$= 197.946/5 = 39.584$$

▪ ان عدد العينات التي يمكن سحبها بحجم 3 هي : $\binom{N}{n} = 10$

▪ ان العينات التي عددها 10 هي كما يلي :

التوزيع الاحتمالي النسبي $f(\bar{x})$	التكرار f_i	ترتيب $\bar{x}$	متوسط العينة $\bar{x}$	قيم وحدات العينة x_i	العينة
0.1	1	2	3	0, 3, 6	A b c
0.2	2	3	2	0, 3, 3	A b d
0.1	1	4	7	18, 3, 0	A b e
0.2	2	7	3	0, 6, 3	A c d
0.2	2	8	8	0, 18, 6	A c e
0.2	2	9	7	0, 18, 3	A d e
=1.0	$\sum f_i = 10$		4	3, 6, 3	B c d
			9	18, 6, 3	B c e
			8	3, 18, 3	B d e
			9	3, 18, 6	C d e

وان :

$$E(\overline{x}) = \sum f_i \overline{x}_i / \sum f_i = \mu$$
$$= 60/10 = 6$$

أي : $E(\overline{x}) = \mu$

$$var(\overline{x}) = \sum f_i (x - \mu)^2 / \sum f_i$$
$$= 65.98/10 = 6.598$$

وباستخدام معامل التصحيح للمجتمع المحدود (N-n)/(N-1) نحصل على :

$$var(\bar{x}) = (\sigma^2/n). (N-n)/(N-1)$$
$$= (39.589/3)\ (3/4)$$
$$= 6.598$$

5-3 تحديد حجم العينة Sample Size Determination

يعد موضوع تحديد عدد وحدات المجتمع التي ينبغي شمولها بالعينة من المسائل المهمة في عملية تصميم العينة، وذلك لتجنب اخذ عينة صغيرة جدا يكون تقديرها للمجتمع غير دقيق وغير مفيد، او سحب عينة كبيرة جدا يؤدي الى زيادة في الجهد والتكاليف الاقتصادية .

وتعتمد عملية تحديد حجم العينة على مقياس تعيين درجة الدقة المستهدفة، وهو ما يعبر عنه بحجم الخطأ المسموح به في ايجاد التقديرات. كأن يكون 0.05 او 0.01...الخ، ويعتمد مقدار هذا الخطأ بطبيعة الحال على طبيعة المجتمع المبحوث ان كان يحتاج الى دقة عالية جدا كما في صناعة الادوية مثلا ويعتمد كذلك على خبرة الباحث، فاذا افترضنا ان الخطأ المسموح به لمتوسط العينة المسحوبة من مجتمع طبيعي هو 0.05 واردنا التاكد من عدم تجاوز هذه النسبة فسيكون لدينا فترة الثقة هي :

$$\overline{x} \mp 1.96 \frac{s}{\sqrt{n}}$$

حيث ان :

$\bar{x}$ متوسط العينة و $\frac{s}{\sqrt{n}}$ تمثل تقدير الخطأ للمعياري للمجتمع، وان s و n هما الانحراف المعياري للعينة وحجمها على التوالي .

3- 5- 1 طريقة حساب حجم العينة

فلو فرضنا ان مقدار الخطا المسموح به والذي هو عبارة عن مقدار الفرق بين متوسطي المجتمع μ والعينة $\bar{x}$ ولنرمز له بـ d، هو :

$$d = |\bar{x} - \mu|$$

من التوزيع الطبيعي نعلم بان :

$$p(|\bar{x} - \mu| \leq z_{(1-\alpha/2)}\sigma_{\bar{x}}) = 1 - \alpha$$

اذن :

$$p(d \leq z_{(1-\alpha/2)}\sigma_{\bar{x}}) = 1 - \alpha$$

وان :

$$(d = z_{(1-\alpha/2)}\sigma_{\bar{x}})$$

$$\left(d = z_{(1-\alpha/2)} \frac{\sigma}{\sqrt{n}}\right)$$

وعليه فان :

$$n = \frac{z^2\sigma^2}{d^2}$$

وبالاعتماد على نتائج العينة في تقدير تباين المجتمع، تصبح صيغة العلاقة :

$$n = \frac{z^2 s^2}{d^2}$$

وفي حالة ان نحصل من تطبيق الصيغة اعلاه على حجم عينة تزيد على 10٪ من حجم المجتمع N فبالامكان تعديله لغرض تخفيضها باستخدام العلاقة التالية: (Snedecar and Cohran, 1980).

$$n' = \frac{n}{1 + \frac{n}{N}}$$

حيث ان n' يمثل حجم العينة بعد التعديل

$$n' = \frac{Nz^2\sigma^2}{Nd^2 + z^2d^2}$$

مثال (9.3) : مجتمع احصائي يتكون من 1000 اسرة، واختير معدل عدد افراد الاسرة كمتغير لتحديد حجم العينة،س وكان الانحراف المعياري في عدد افراد الاسرة للمجتمع هو $\sigma = 5$، والمطلوب تحديد حجم العينة المطلوب سحبها عند درجة ثقة مقدارها 95 ٪، علما بان الفرق بين متوسطي العينة والمجتمع هو $d = 1$.

الحل لـ (9.3) : لدينا :

$$N = 1000,$$
$$\sigma^2 = 25,$$
$$d = 1,$$
$$z_{1-\alpha/2} = z_{0.975} = 1.96$$

باستخدام الصيغة التالية نحصل :

$$n' = \frac{n}{1 + \frac{n}{N}} = \frac{Nz^2\sigma^2}{Nd^2 + z^2d^2}$$

$$= \frac{(1000)(3.8416)(25)}{1000(1) + (3.8416)(1)} = \frac{96040}{1003.8416} \approx 96$$

وهو حجم العينة المطلوب شمولها لدراسة متوسط عدد افراد الاسرة .

مثال (10.3): لتوفير معلومات لتحديد حجم العينة فقد اختيرت عينة تجريبية لتقدير متوسط محصول القمح للحيازة الزراعية الواحدة في منطقة ما، ووجد ان مقدار التباين في المحصول في هذه العينة هو $s^2 = 90.3$ كغم. فما هو عدد الحيازات المطلوب شمولها بالمسح الاحصائي بحيث لايزيد مقدار الفرق بين متوسطي العينة والمجتمع عن $d = 1.5$ كغم وبدرجة ثقة مقدارها ٪ 95 .

الحل لـ (10.3) : لدينا

$$S^2\ 90.3 \quad ، d^2 = 2.25$$

$$z_{1-\alpha/2} = z_{0.975} = 1.96$$

$$n = \frac{z^2\sigma^2}{d^2} = \frac{(3.8416)(90.3)}{2.25} = 154$$

وهو عدد الحيازات المطلوب شمولها كعينة في المسح الاحصائي .

3- 5- 2 طريقة حساب حجم العينة في حالة النسب

وهي الحالات التي يعبر فيها عن خاصية المجتمع بنسب كنسبة المتزوجين في المجتمع او نسبة الوحدات الصالحة في الانتاج، والتي يطلق عليها بنسبة النجاح والفشل. فنفترض بان توزيع هذه المجتمعات هي مقاربة للتوزيع الطبيعي، وعليه يتم الاستعاضة عن σ^2 بالمقدار PQ وان $d = z\sqrt{v_{ar}(p)}$.

وان صيغة حساب حجم العينة هي كما يلي :

(1) في حالة معلومية نسبة خاصية المجتمع P وحجمه N فان صيغة الحساب تكون :

$$n = \frac{NZ^2(PQ)}{Nd^2 + Z^2(PQ)}$$

حيث ان: $Q = 1 - P$

مثال (11.3) : يقوم مصنع لصناعة الالبان بانتاج 10000 وحدة من الاجبان المتعدد الانواع، وفي المعدل فان 10٪ من وحداته المنتجة لاتطابق المواصفات المقررة، فما هو حجم العينة المطلوب سحبها لتقدير نسبة الوحدات غير المطابقة للمواصفات، بحيث لايتجاوز الفرق في تقدير النسبة عن 0.02 وبدرجة ثقة مقدارها 90 ٪ .

الحل لـ (11.3) : لدينا

$N = 10000$ ، $P = 0.10$ ، $Q = 0.90$ ، $d = 0.02$ ، $Z_{(1-\alpha/2)} = Z_{0.95} = 1.64$

وباستخدام الصيغة نحصل على :

$$n = \frac{NZ^2(PQ)}{Nd^2 + Z^2(PQ)} = \frac{2420.64}{4.242} = 571$$

وهي عدد الوحدات العينة المطلوبة من منتجات الالبان .

(2) اما في حالة معلـــومية حجم المجتمع فقط مع مجهولية نسبة الخاصية، فيستعاض بنسبـة العينة p بدلا من نسبة المجتمع، وتصبح صيغة الحساب هي:

$$n = \frac{NZ^2pq}{Nd^2 + Z^2pq}$$

(3) في حالة عدم معلومية حجم المجتمع N ونسبة الخاصية ايضا مع معلومية تقارب المجتمع للتوزيع الطبيعي فيتم استخدام الصيغة التالية :

$$n = \frac{Z^2pq}{d^2}$$

مثال (12.3) : قام احد اصحاب مشاتل النباتات الظلية بفحص عينة تجريبية تتكون من 480 نبته، فوجد 15٪ منها مصابة بمرض. فما هو حجم العينة التي يستطيع بضوئها تحديد نسبة الشتلات المصابة، ضمن فرق مقداره 0.05 بين متوسطي العينة والمجتمع، وعند درجة ثقة مقدارها 95 ٪ .

الحل لـ (12.3) : لدينا

$P = 0.15$ ، $q = 0.85$ ، $d = 0.05$ ، $Z_{0.975} = 1.96$

وبتطبيق الصيغة نحصل على :

$$n = \frac{Z^2pq}{d^2} = \frac{0.4898}{0.0025} = 196$$

اي ان صاحب المشتل يحتاج لفحص 196 نبته كعينة لمعرفة نسبة النبتات المصابة.

3- 5- 3 طريقة حساب حجم العينة وفق الامكانات المتاحة

وتستخدم في الحالات التي تكون فيها الامكانيات المالية محدودة، ويمكن التذكير هنا بانه في العديد من الدول يقوم الباحث او الجهة القائمة بالبحث بدفع مبالغ مالية احيانا او كلف تدريبيه او على شكل هدايا للمبحوث (وحدة العينة) مقابل ما يقدمه من معطيات لانه قد يتطلب في بعض الدراسات الاستعانة باجهزة ومعدات لاغراض طبية او هندسية وما شابه التي تتطلب من المبحوث التفرغ او المكوث لاوقات محددة، او تتطلب الحالة تكرار العودة بصورة دورية الى نفس وحدة العينة كأن تكون الاسرة لاغراض متابعة تغييرات اسعار الخدمات او تطور مستوى المعيشة او مراقبة ظواهر اقتصادية واجتماعية محددة، عندها يمكن توظيف العلاقة التالية في حساب حجم العينة:

$$C = c_0 + n\,c_1$$

حيث ان :

C هي الميزانية المالية للاستقصاء (المسح)

c_0 النفقات العامة للاستقصاء (التي لاتدخل ضمن فقرة c_1)

c_1 كلفة ملئ الاستبانة الواحدة واستخراج نتائجها .

6-3 العينة العشوائية الطبقية Stratified Random Sample

3- 6- 1 مفهوم العينة واستخداماتها

عندما يكون المجتمع الاحصائي غير متجانس، تصبح العينة العشوائية البسيطة غير مناسبة للاستخدام لانها سوف لاتكون ممثلة للمجتمع الذي تسحب منه، لذا يتطلب الامر اللجوء الى العينة العشوائية الطبقية التي تتلخص اختيار وحداتها بما يلي :

(1) تقسيم المجتمع الاحصائي N غير المتجانس الى مجتمعات صغيرة متجانسة: N_1، N_2، .... N_k ، وتسمية هذه المجتمعات بالطبقات Strata على ان لايحصل تداخل بين وحداتها، اي لا تتكرر الوحدة ذاتها في اكثر من طبقة واحدة، بحيث يتحقق $N_1+N_2+....+N_k = N$

(2) نختار عينة عشوائية بسيطة من كل طبقة، بحيث تكون العينات المختارة من الطبقات المختلفة هي العينة العشوائية الطبقية، اي $n1+n2+.....+nk = n$

3- 6- 2 طريقة تحديد عدد وحدات العينة لكل طبقة

والمقصود هنا هو كيفية تحديد حجم العينة العشوائية البسيطة التي يتم سحبها من كل طبقة، ونتطرق هنا الى طريقتين رئيسيتين هما :

(1) طريقة الاختيار المتناسب Proportional allocation method

وبموجب هذه الطريقة فان حجم العينة لكل طبقة يكون متناسبا مع نسبة حجم الطبقة الى الحجم الكلي للمجتمع الاحصائي، اي ان حجم العينة العشوائية الماخوذة من طبقة ما الى حجم العينة النهائي يكون مساويا لنسبة حجم تلك الطبقة الى الحجم الكلي للمجتمع، ويمكن التعبير عن ذلك بالصيغة التالية :

$$W_i = \frac{N_i}{N} = \frac{n_i}{n}$$

حيث ان W_i هي نسبة العينة i الى الحجم الكلي للعينة، بهذا يكون حجم العينة i من الطبقة i هو :

$$n_i = n\frac{N_i}{N}$$

حيث ان :

n حجم العينة الكلي، أي $\Sigma n_i = n$ و N حجم المجتمع الكلي، اي $\Sigma N_i = N$

مثال (13.3): لنفترض ان لدينا مجتمعا يتكون من 25 اسرة وان المصروفات النثرية الاسبوعية (بالدولار) لكل من هذه الاسر هو كما مبين في الاتي، والمطلوب سحب عينة عشوائية طبقية تتكون من 8 اسر مستخدما طريقة الاختيار المتناسب .

10، 50، 40، 15، 41، 24، 23، 25، 45،48، 18، 17، 27، 30، 38، 32، 12، 14، 16، 19، 44، 43، 42، 46، 29

الحل لـ (13.3) : من ملاحظة ارقام المجتمع الاحصائي نستدل على امكانية تقسيم المجتمع الى ثلاث طبقات هي :

الطبقة 1 (N_1) : 10، 15، 17، 12، 18، 14، 16، 19

الطبقة 2 (N_2) : 32، 23، 24، 25، 27، 29، 30

الطبقة 3 (N_3) : 50، 40، 41، 45، 38، 42، 46، 44، 43، 48

اي ان عدد وحدات كل طبقة هو : $N_1=8$، $N_2=7$، $N_3=10$

وباستخدام صيغة تحديد عدد الاسر المطلوب سحبها من عشوائيا من كل طبقة نحصل على:

$$n_i = n\frac{N_i}{N}$$

وهي عدد وحدات عينة الطبقة N_1 : $n_1 = 8\frac{8}{25} = 2.56 \approx 3$

وهي عدد وحدات عينة الطبقة N_2 : $n_2 = 8\frac{7}{25} = 2.24 \approx 2$

وهي عدد وحدات عينة الطبقة N_3 : $n_3 = \frac{10}{25} = 3.2 \approx 3$

وفي المرحلة الاخيرة نستخدم الجداول العشوائية على وفق الخطوات المذكورة في (4) من الفقرة (3.3) نحصل على وحدات العينة التي ظهرت من كل طبقة على النحو الاتي :

العينة n_1 : 10، 17، 14

العينة n_2 : 27، 23

العينة n_3 : 38، 41، 44

وبذلك فان وحدات العينة هي: 38، 41، 44، 27، 23، 10، 14، 17

(2) طريقة الاختيار الامثل Optimal allocation method

وتقوم هذه الطريقة على اساس تقليل التباين او التكاليف الى الحد الادنى عند تحديد احدهما، فان عدد وحدات كل طبقة سيتناسب مع درجة تجانس وحداتها فيكون صغيرا في حالة الطبقات المتجانسة في حين يزداد في حالة الطبقات غير المتجانسة، اي ان تحديد العدد يعتمد على مقدار تباين مجتمع كل طبقة بالاضافة الى حجم الطبقة ذاتها، وتدعى هذه العلاقة بالاختيار الامثل لنيمان (Nymen)، ويمكن التعبير عن هذه العلاقة في حالة عدم تساوي تكاليف اختيار الوحدة بالصيغة التالية :

$$n_i = n\frac{\frac{N_i S_i}{\sqrt{C_i}}}{\frac{\sum_{i=1}^{k} N_i S_i}{\sqrt{C_i}}}$$

حيث ان :

C_i هي تكاليف اختيار الوحدة n_i، k عدد الطبقات، وان صيغة دالة التكاليف الخطية هي :

$$C = C_o + \sum_{i=1}^{k} C_1 n$$

أما اذا كانت تكاليف اختيار الوحدة متساوية فتصبح العلاقة على النحو الاتي:

$$n_i = n\frac{N_i S_i}{\sum_{i=1}^{k} N_i S_i}$$

حيث ان n هي حجم العينة الطبقية، و S_i هو الانحراف المعياري للطبقة i وهو

$$S_i = \sqrt{\frac{\sum\left(X_i - \overline{X}\right)^2}{N_i}}$$

وان المقدار: $\frac{N_i S_i}{\sum_{i=1}^{k} N_i S_i}$ يمثل النسبة W_i في حالة طريقة الاختيار المتناسب.

مثال (14.3): من دراسة سابقة شملت خصائص المسافرين (العمر، الدخل، هدف السفر... الخ) اتضح بان عدد المسافرين على مجموعة خطوط السير التي تربط عاصمة احدى الدول بالمدن الرئيسية الاخرى لتلك الدولة هو 70967 مسافرا اسبوعيا وان توزيعهم حسب ايام الاسبوع موضح في الجدول (1.3) الاتي، ولغرض دراسة تقوم بها هيئة التخطيط الحضري، فقد تم تحديد حجم العينة بـ 300 مسافر لشمولهم بهذه الدراسة وذلك بالاعتماد على متغيرمعدل الدخل الشهري لهذا المجتمع، وتم تقسيم المجتمع الى 7 طبقات (ايام الاسبوع) تبعا لمتغير الدخل وان حجم كل طبقة N_i وانحرافها المعياري S_i (وفقا لمخرجات SPSS بسبب كبر حجم المجتمع) هي مبينة في الجدول المذكور. والمطلوب تحديد عدد الوحدات الازم سحبها من كل طبقة باستخدام طريقة الاختيار الامثل .

جدول رقم (1.3)

عدد المسافرين من عاصمة احدى الدول اسبوعيا حسب ايام الاسبوع

$N_i S_i$	الانحراف المعياري S_i	حجم الطبقة N_i	الطبقة i
31265.9	3.502	8928	N_1
38796.4	4.527	8570	N_2
30950.8	3.596	8607	N_3
28870.8	3.245	8897	N_4
34384.0	3.500	9824	N_5
43821.5	3.444	12724	N_6
72572.6	5.409	13417	N_7
$\sum_{i=1}^{7} N_i S_i = 280662$		**70965**	**المجموع**

الحل لـ (14.3):

بتطبيق صيغة طريقة الاختيار الامثل لحالة التكاليف المتساوية، فان عدد الوحدات المطلوب سحبها من كل طبقة Ni هو :

$$n_i = n\frac{N_i S_i}{\sum_{i=1}^{k} N_i S_i}$$

$$n_1 = 300\frac{(8928)(3.502)}{280662} = 33$$

$$n_2 = 300\frac{(8570)(4.527)}{280662} = 41$$

$$n_3 = 300\frac{(8607)(3.596)}{280662} = 33$$

$$n_4 = 300\frac{(8897)(3.245)}{280662} = 31$$

$$n_5 = 300\frac{(9824)(3.500)}{280662} = 37$$

$$n_6 = 300\frac{(12724)(3.444)}{280662} = 47$$

$$n_7 = 300\frac{(13417)(5.409)}{280662} = 78$$

ولكبر حجم العينة يتعذر استخدام الجداول العشوائية في تحديد الوحدات التي يتم سحبها من كل طبقة، لذلك نستعين بالحاسوب لمثل هذه الحالة، سواء بواسطة برنامج SPSS او برنامج EXCEL توفيرا للوقت والجهد.

3-6-3 تقدير متوسط المجتمع وتباينه للعينة العشواييه الطبقية

(1) تقدير متوسط المجتمع

يكون مقدر متوسط العينة الطبقية الذي نرمز له بـ $\overline{X}_{st}$ هو :

$$\overline{X}_{st} = \frac{\sum_{i=1}^{k} N_i \overline{X}_i}{N}$$

$$= W_i \overline{Xi}$$

حيث ان :

$$\overline{X}_i = \frac{\sum_{i=1}^{ni} X_i}{n_i}$$

(2) تقدير تباين متوسط المجتمع

ان مقدر تباين متوسط العينة الطبقية هو :

$$v_{st}\left(\overline{X}_{st}\right) = \frac{1}{N^2} \sum N_i \left(N_i - n_i\right) \frac{S_i}{n_i}$$

$$\textbf{Or :} \quad \sum W_i \frac{S_i^{\ 2}}{n_i} \left(1 - \frac{n_i}{N_i}\right)$$

حيث ان :

$$S_i^{\ 2} = \frac{\sum_{i=1}^{ni} \left(X_i - \overline{X}_I\right)^2}{n-1}$$

وان :

$$S_{st} = \sqrt{v_{st}(\overline{X}_{st})}$$

7-3 العينة العشوائية المنتظمة Systematic Random Sample

3-7-1 مفهوم العينة واستخداماتها

لقد لاحظنا بان العينات التي تطرقنا اليها وهي العشوائية البسيطة والعشوائية الطبقية كانت تتطلب معرفة حجم المجتمع وغالبا ما تكونا

مكلفتين، واحيانا يكون استخدامهما مستحيلا لعدم معرفة حجم المجتمع، ولحل مثل هذه المشكلات برزت طريقة المعاينة العشوائية المنتظمة والتي تتلخص في اختيار i^{th} على التوالي بعد تحديد نقطة البداية عشوائيا بين الاعداد من 1 ،2، ...... ،i وقد سميت بالعينة العشوائية المنتظمة، لان وحداتها يتم اختيارها بطريقة منتظمة بعد نقطة البداية العشوائية.

فمثلا اذا اردنا اختيار عينة عشوائية منتظمة، باختيار كل عاشر وحدة، فان علينا ان نحدد نقطة البداية عشوائيا من بين 1 و 10 وليكن 4 حينئذ تكون وحدات العينة المنتظمة هي : 4، 14، 24، 34 .... الخ والى ان نحصل على عدد وحدات العينة المطلوبة. والعينة العشوائية المنتظمة واسعة الاستخدام وخاصة في حالة المجتمعات المتحركة كوسائط النقل المارة او حركة المسافرين وما شابه وكلك في المجالات التطبيقية كالمترددين على المكتبات العامة او المتسوقين من المخازن التجارية او اختيار عينة من المساكن او المتاجر وهكذا. ويتميز هذا النوع من العينات بانخفاض تكاليفه و بسهولة التطبيق حيث كل مانحتاجه هو تحديد عدد عشوائي واحد، اضافة الى انها تتوزع على المجتمع توزيعا منتظما اكثر مما يحصل مع باقي العينات التي قد تتركز وحداتها في موقع واحد .

3- 7- 2 اسلوب اختيار وحدات العينة العشوائية المنتظمة

في حالة معرفة حجم المجتمع N فان اختيار عينة عشوائية منتظمة بحجم n يتم على النحو الاتي :

(1) نحدد طول دورة المعاينة L وهي: $L = \frac{N}{n}$

(2) نحدد نقطة البداية باختيار عدد عشوائيا بين 1 و L

(3) نضيف في كل مرة طول الدورة L الى العدد الذي تم اختياره لنحصل على حجم العينة n المطلوب، فاذا اردنا مثلا اختيار عينة عشوائية منتظمة بحجم n = 10 من مجتمع مكون من 100 وحدة يستوجب اتباع الخطوات المذكورة وكالاتي :

- نجد طول الدورة وهي: $L = \frac{100}{10} = 10$
- نحدد نقطة البداية، اي الوحدة الاولى بالعينة وذلك عشوائيا من بين الاعداد التي تقع بين 1 و 10 وليكن 4
- نحدد عناصر العينة باضافة طول الدورة 10 الى العدد الاول 4 بانتظام فنحصل على وحدات العينة وهي: 4، 14، 24، 34،44، 54، 64، 74،84، 94، اي :

4 ،4+2L ،2+3L ،.................. ،4+ (n-1)L

الا ان للعينة العشوائية المنتظمة عيبان، احدهما حاصل والثاني محتمل الوقوع. فالعيب الحاصل يتمثل في انه لايوجد للعينة العشوائية المنتظمة طريقة ذات اعتمادية عالية في تقدير الخطأ المعياري لمتوسط المجتمع، فرغم شمولها ضمنيا على طبقات الا ان العشوائية تحصل مع مفردة واحدة لكل طبقة. اما العيب المحتمل الوقوع فيحصل عندما تاخذ وحدات المجتمع نسقا دوريا ثابتا، فمثلا عند الرجوع الى ترتيب افراد الاسرة يبدا عادة برب الاسرة ومن ثم الزوجة فالاولاد الاكبر فالاصغر وهكذا، ففي مثل هذه الحالة تكون الوحدة الاولى دائما رب الاسرة والثانية غالبا الزوجة والثالثة الابن الاكبر وكذا. وعليه اذا كان ترتيب وحدات المجتمع موضوع الدراسة ترتيبا دوريا فيجب تجنب استخدام هذا النوع من العينات .

3- 7- 3 تقدير الخطأ المعياري للعينة وتقدير مجموع المجتمع

ويتم فيها استبدال الانحراف المعياري للمجتمع σ بالانحراف المعياري للعينة S بموجب الصيغة التالية :

$$S_{sy} = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$$

حيث ان :

$$S_{sy} = \sqrt{v_{ar}(\bar{x}_{sy})}$$

وان حساب مجموع المجتمع التقديري لاي متغير يتم بضرب متوسط العينة المنتظمة $\bar{x}_{sy}$ بحجم المجتمع N .

8-3 العينة العشوائية العنقودية Cluster Random Sample

3- 8- 1 مفهوم العينة واستخداماتها

وتستخدم في الحالات التي تكون فيها وحدات المجتمع على شكل تجمعات متشابهه الى حد كبير بالنسبة للخاصية التي نقوم بدراستها مثل المدن، القرى، الشوارع، الكليات، الاسواق، المناطق الزراعية وغيرها، وتدعى هذه التجمعات بالعناقيد، ويحتوي كل عنقود على عدد من عناصر المجتمع الاصلية التي غالبا ما تكون متجانسة. وياتي استخدام هذا النوع من العينات لسببين هما :

(1) حين لايتوفر اطار احصائي دقيق للمجتمع، وتكون كلفة تهيئة مثل هذا الاطار عالية، فلو افترضنا اننا بصدد اجراء استقصاء (مسح) اقتصادي واجتماعي وكانت وحدة المشاهدة فيه هي الاسرة، لكن لاتتوفر قوائم باسماء الاسر بينما تتوفر قوائم باسماء المناطق او الاحياء المتشابهه في الخاصية التي نقوم بدراستها (كالدخل او مستوى المعيشة ...الخ)، حيث تضم كل منطقة او حي مجموعة اسر، ففي مثل

هذه الحالة يمكن اختيار عينة عشوائية من الاحياء ثم دراسة اسر هذه الاحياء المختارة او عينة من كل مها تبعا لطبيعة الدراسة .

(2) من اجل تركيز الجهود والاموال المرصودة في مجال معين الذي من شانه تقليل الكلفة والجهد والوصول الى وحدات المجتمع .

3- 8- 2 اسلوب اختيار العينة العشوائية العنقودية

هناك عدة طرق لاختيار لاختيار العينة العشوائية العنقودية وهي: اما ان تتم عملية الاختيار على مرحلة واحدة يجري فيها اختيار عينة عشوائية بسيطة من العناقيد ثم دراسة وحدات هذه العتاقيد، او اختيارها على اكثر من مرحلة، نقوم في المرحلة الاولى باختيار عينة عشوائية بسيطة من العناقيد ثم يتم بعدها اختيار عينة عشوائية بسيطة من كل عنقود مختار تم اختياره في المرحلة الاولى لتكون العينة العشوائية العنقودية، وبذلك يكون الاختيار قد تم على مرحلتين وهكذا .

3- 8- 3 تقدير متوسط المجتمع وتباينه من العينة العشوائية العنقودية

نقتصر هنا على تقديرات العينة العنقودية لمرحلة واحدة، ونتطرق في العينة العشوائية متعددة المراحل عن حالة التقديرات في حالة مرحلتين او اكثر.

لنفترض لدينا M من العناقيد، يحتوي كل منها على N_i من من الوحدات، وان عدد الوحدات في المجتمع هو $\sum N_i = N$ وسحبث عينة عشوائية بسيطة من m عنقود، ورمزنا لقيمة الوحدة j بالعنقود i بالرمز x_{ij} ولمجموع العنقود i بالرمز x_iفان متوسط العينة العنقودية ولنرمز له بـ x_c سيكون على النحو الاتي :

$$\overline{x_c} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{Ni} x_i}{\sum N_i} = \frac{\sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} N_i}$$

وهو مقدر لمتوسط المجتمع

اما في حالة عدم معلومية N فيكون مساويا الى :

$$M_{\overline{xi}} = \frac{M\sum x_i}{m}$$

اما مقدار تباين المتوسط $\overline{x_c}$ فيتم ايجاده باستخدام العلاقة :

$$v_{ar}\left(\overline{x_c}\right) = S_c^{\,2} \cdot \frac{(M-m)}{Mm\left(\overline{M}\right)^2} \cdot \frac{\sum_{i=1}^{m}\left(x_i - \overline{x_c}N_i\right)^2}{(m-1)}$$

أو

$$= \frac{M(M-m)}{mN^2} \cdot \frac{\sum_{i=1}^{m}\left(x_i - \overline{x_c}N_i\right)^2}{m-1}$$

حيث ان M هو متوسط حجم العنقود في المجتمع N وفي حالة عدم معرفة M يمكن تقديره بمتوسط حجم العنقود بالعينة، اي :

$$\overline{m} = \frac{\sum_{i=1}^{m} N_i}{m}$$

ويكون مقدر المجموع الكلي في المجتمع في حالة معرفة N :

$$N\overline{x_c} = \frac{N\sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} N_i}$$

وفي حالة عدم معلومية N فيكون المجموع الكلي مساويا الى :

$$M\overline{x_c} = \frac{M\sum x_i}{m}$$

مثال (15.3) : اختيرت عينة عشوائية بسيطة من 4 عناقيد من مجتمع يحتوي على 20 عنقود وحصلنا على المعطيات التالية، والمطلوب ايجاد تقديرات كل من متوسط وتباين ومجموع المجتمع .

العنقود i	حجم العنقود N_i	مجموع قيم العنقود x_i
1	8	96
2	4	42
3	6	52
4	5	54
المجموع	$\sum N_i = 23$	$\sum x_i = 244$

الحل لـ (15.3) :

- نستخدم صيغة تقدير متوسط المجتمع وهي :

$$\overline{x}_c = \frac{\sum_{i=1}^{m}\sum_{j=1}^{Ni} x_i}{\sum N_i} = \frac{\sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} N_i}$$

$$= \frac{244}{23} = 10.61$$

- نستخدم صيغة حساب تقدير تباين المجتمع وذلك باستخدام $\overline{m}$ بدلا من $\overline{M}$ من دون الاعتماد على حجم المجتمع N فنحصل على:

$$S_{\overline{x}_c}^{\ 2} = \frac{(M-m)}{Mm(\overline{m})^2} \cdot \frac{\sum_{i=1}^{m}\left(x_i - \overline{x}_c N_i\right)^2}{m-1}$$

$$= \frac{(M-m)}{Mm(\overline{m})^2} \cdot \frac{\sum\left(x_i^{\ 2} - 2x_i \overline{x}_c N_i + \overline{x}_c N_i^{\ 2}\right)}{m-1}$$

$$= \frac{(M-m)}{Mm(\overline{m})^2} \cdot \frac{\left(\sum x^2{}_i - 2\overline{x_c}\sum x_i N_i + \overline{x_c}^2 \sum N_i^2\right)}{m-1}$$

$$S_{\overline{x_{c_c}}}^2 = \frac{(20-4)}{(20)(4)\left(\frac{23}{4}\right)^2} \times \frac{\left[16600-(2)(10.61)(1518)+(141)(10.61)^2\right]}{4-1}$$

$$= 1.58$$

- اما لتقدير المجموع الكلي للمجتمع $M\overline{x_i}$ وذلك لعدم معلومية حجم المجتمع N .

$$\overline{x_i} = \frac{\sum x_i}{m} = \frac{244}{4} = 61$$

وعليه فان: $M\overline{x_i} = (20)(61) = 1220$

3-9 العينة العشوائية المتعددة المراحل

Multi-Stage Random Sample

3- 9- 1 مفهوم العينة واستخداماتها

ذكرنا في الفقرة السابقة ان العينة العنقودية قد تتم على مرحلة واحدة او على اكثر من مرحلة واحدة، ففي العديد من الدراسات التطبيقية نجد ان وحدات المجتمع الاصلي الذي نريد دراسته توجد على شكل تجمعات كبيرة (اولية)، وان كل تجمع يحتوي على عدة تجمعات (ثانوية)، وكل تجمع من هذه التجمعات الثانوية يحتوي على تجمعات اخرى وهكذا. وفي ضوء ذلك يصعب الحصول على اطار للمجتمع او يكون مكلفا، وعليه يتم اللجوء الى عينة عشوائية من التجمعات الاولية وهو ما يسمى بالمحلة الاولى، ثم نختار عينة عشوائية من كل تجمع تم اختياره في المرحلة الاولى وهو مايسمى بالمرحلة الثانية، بعدها نختار عينة عشوائية من كل تجمع تم اختياره

في المرحلة الثانية وهو ما يسمى بالمرحلة الثالثة وهكذا، وبذلك يتم اختيار العينة العشوائية النهائية على مراحل عدة، وتدعى هذه العينة بالعينة العشوائية متعددة المراحل.

قد تكون التجمعات في كل مرحلة متجانسة او غير متجانسة، وعليه نختار عينة عشوائية بسيطة في كل مرحلة اذا كانت التجمعات متجانسة في الصفة التي نقوم بدراستها، في حين يتم اللجوء الى العينة العشوائية الطبقية اذا كانت التجمعات غير متجانسة. فمثلا اذا كان لدينا مجتمع يتكون من ارياف متشابهه في الظاهرة التي نقوم بدراستها، نقوم في المرحلة الاولى باختيار عينة عشوائية بسيطة من هذه الارياف، وان كل ريف يتكون من اخرى ولتكن عدد من المدارس المتجانسة في الصفة، عليه نقوم باختيار عينة عشوائية بسيطة من كل تجمع تم اختياره في المرحلة الاولى لنحصل على تجمعات المرحلة الثانية، وقد تكون تجمعات المرحلة الثانية مكونة ايضا من تجمعات متجانسة من الطلبة، حينئذ نقوم باختيار نقوم باختيار عينة عشوائية بسيطة من التجمعات التي اختيرت في المرحلة الثانية لنحصل على العينة المطلوبة في المرحلة الثالثة. ومن هنا جاء تسميتها بالعينة المتعددة المراحل، ونتيجة للتجانس الذي تؤول اليه وحدات العينة فان نتائجها تكون قريبة من واقع المجتمع الاحصائي .

ولتقريب الصورة نرمز للتجمعات التي يتكون منها المجتمع بـ M، وكل من هذه التجمعات يضم N_i من الوحدات، اي: $i = 1$، 2، 3، .... M، ثم اختيرت عينة عشوائية بسيطة ولنرمز لها بـ m وهي المرحلة الاولى، ومن تجمعات m تم اختيار عينة عشوائية بسيطة بحجم n_i من كل من تجمعات المرحلة الاولى لتكون المرحلة الثانية، وبذلك نكون قد حصلنا على العينة

العشوائية البسيطة على مرحلتين وكما هو موضح في الشكل البياني رقم (3.3) التالي :

شكل بياني رقم (3.3)

يبين مخططا توضيحيا لاختيار عينة عشوائية بسيطة على مرحلتين

التجمع	1	2	3	M
حجم التجمع	N_1	N_2	N_3	N_m
متوسط التجمع	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$	$\overline{X}_m$
المرحلة الاولى (التجمعات المختارة)	N_1	N_2	N_3	N_m
المرحلة الثانية (العينات المختارة)	n_1	n_2	n_3	n_m
متوسطات العينات	$\overline{x}_1$	$\overline{x}_1$	$\overline{x}_1$	$\overline{x}_m$

3-9-2 تقدير متوسط المجتمع وتباينه من العينة العشوائية متعددة المراحل

(1) تقدير متوسط المجتمع في حالة معلومية حجم المجتمع N هو :

$$\overline{x}_{mn} = \frac{M}{N} \cdot \frac{\sum_{i=1}^{m} N_i \overline{x_i}}{m}$$

(2) تقدير متوسط المجتمع في حالة مجهولية حجم المجتمع N هو:

$$\overline{x}_{mn} = \frac{\sum_{i=1}^{m} N_i \overline{x_i}}{\sum N_i}$$

(3) تقدير تباين متوسط المجتمع $\overline{x}_{ms}$

وللحاجة الى ايجاد التباين بين التجمعات وكذلك ضمن المجاميع لأجل تقدير تباين متوسط المجتمع، يفضل الاستعانة بتحليل التباين لهذا الغرض، وهو

موضوع الفصل الثامن من هذا الكتاب. اما عندما يكون حجم العناقيد متساويا وحجم العينات المختارة من كل عنقود متساوية ايضا وباهمال معامل التصحح،فيمكن استخدام العلاقة التالية في حالة اختيار العينة على مرحلتين:

$$v_{ar}(\overline{X}_{mn}) = \frac{S_1^{\ 2}}{n_1} + \frac{S_2^{\ 2}}{n_1 n_2}$$

حيث ان S_1^2، S_2^2، n_1، n_2 هي تباين وحجم عينة المرحلة الاولى والمرحلة الثانية على التوالي .

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i}^{n_1} (x_i - \overline{x})^2$$

هو مقدار التباين بين العينات

$$S_2^2 = \frac{1}{n_1(n_2 - 1)} \sum_{i}^{n_1} \sum_{j}^{n_2} (x_{ij} - \overline{x}_i)^2$$

وهو مقدار التباين ضمن (داخل) العينات

3- 10 حالات دراسية في العينات العشوائية

3- 10- 1 حالة دراسية رقم C_{3-2} :

في تصميم عينة مسح لنشاط خدمات الوساطة المالية

(تعدد الانشطة مع معلومية المعالم الرئيسية)

في مسح احصائي عن الخدمات والوساطة المالية في إمارة أبو ظبي استهدف توفير المعطيات الإحصائية المتعلقة بحجم الإنتاج ومستلزماته، وحجم رؤوس الأموال ومصادرها، وحركة الموجودات وتكوين رأس المال، والوقوف على المشاكل والعقبات التي تواجه قطاع الخدمات والوساطة المالية.

شمل المسح الأنشطة الاقتصادية المتعلقة بالمطاعم والفنادق والأنشطة العقارية والإيجارية وأنشطة المشاريع التجارية وأنشطة البحث والتطوير والخدمات المتعلقة بالحاسب الآلي وخدمات الأعمال القانونية والمحاسبية والاستشارات الإدارية والتجارية والهندسية وخدمات الصحة والعمل الاجتماعي والتعليم وأنشطة الخدمات المجتمعية والاجتماعية والشخصية والوساطة المالية من مصارف وتأمين وتمويل ومحلات صرافة وإدارة الأسواق المالية ومكاتب بيع وشراء الأسهم والأنشطة المساعدة لها.

وكان الإطار الإحصائي الذي تم اعتماده عبارة عن قائمة تضم وحدات المجتمع الإحصائي التي تمارس نشاطها الاقتصادي ممثلة بعدد المنشآت موزعة حسب فئات عدد المشتغلين وهي: فئة أقل من 5 مشتغلين، 5-9 مشتغلين، وفئة 10 فأكثر. وتضم على التوالي نسبة 66٪ و 20٪ و 14٪ من إجمالي عدد المنشآت البالغ 7128 منشأة. ويعود 70٪ من المجموع على منطقة أبو ظبي، والباقي لمنطقة العين. وأن 80٪ من الإجمالي يقع في الحضر.

تضمنت استراتيجية المسح استخدام الشمول الكلي لكل من : المنشآت التي تقع ضمن 10 مشتغلين فأكثر وذلك لأهميتها وقلة حجمها مقارنة بالفئات الأخرى، والمنشآت العاملة في أنشطة التعليم والصحة والعمل الاجتماعي والوساطة المالية لقلة حجمها أيضاً. وتشكل نسبة هذه المنشآت المشمولة كلياً 23٪ من العدد الإجمالي. أما اسلوب العينة فقد تم استخدامه مع المنشآت التي يعمل بها أقل من 5 مشتغلين، وتلك التي يعمل بها 5-9 مشتغلين لكل من أنشطة الفنادق والمطاعم والأنشطة الاجتماعية والمجتمعية والشخصية الأخرى. أي أن نسبة المجتمع الذي تسحب منه العينة يشكل 77٪ من إجمالي المجتمع تحت الدراسة.

(1) ماذا نستدل من معطيات المجتمع تحت الدراسة

- عدم تجانس وحداته وفقاً لأبرز متغيراته (فئات عدد المشتغلين، طبيعة النشاط، حجم رأس المال... الخ).
- مفردات المجتمع موزعة على أكثر من منطقة وأكثر من بيئة وأكثر من نشاط، وكل نشاط على أكثر من مستوى وأكثر من فئة لعدد المشتغلين.
- إمكانية لتوفير إطار المجتمع الإحصائي المطلوب لتصميم العينة.

عليه فإن المعطيات اعلاه تشير إلى إمكانية استخدام العينة العشوائية الطبقية.

(2) تحديد حجم العينة

- **القرار الأول**: هل يتم تحديد حجم العينة مسبقا أم لاحقا، وحيث ان : المجتمع تحت الدراسة معلوم مما يساعد في توفير معالم المجتمع ،لذلك فإن حجم العينة يمكن ان يخضع للتحديد المسبق .
- **القرار الثاني**: صيغة احتساب حجم العينة

حيث إن: متغيرات المجتمع المعتمدة هي كمية. فالصيغة الأكثر كفاءة في التعامل مع هذه الحالة هو الاستعانة بصيغة التوزيع الطبيعي سواء أكان توزيع المجتمع موزع طبيعياً أو مقارب للتوزيع الطبيعي وشكلها العام هو:

$$n' = (NZ^2\sigma^2)/(Nd^2 + Z^2\sigma^2)$$

- القرار الثالث: القيم المناسبة لكل من Z و d، وللوصول إلى القرار المناسب، نحتاج التأمل بالعوامل الحاسمة التالية :
  - هدف وطبيعية مجتمع الدراسة.
  - منهجية وطرق التحليل الإحصائي.

- الإمكانيات المالية والبشرية : $n = (C-C_0)/C_1$

- احتساب حجم العينة : ويتم ذلك باعتماد متغير أساسي للظاهرة التي يتم دراستها وإن الحصول على ذلك يكون بالاعتماد أما على مسوحات سابقة، أو في حالة عدم توفرها الركون على مسح تجريبي. ونتيجة لتوفير نتائج تعدادالمنشآت لعام 1995، فقد تم إيجاد معلمتي الوسط الحسابي μ والانحراف المعياري σ لمتغير عدد المشتغلين باعتباره أساسياً لأي نشاط من جهة ولعدم توفر متغير آخر أكثر أهمية. وإن حساب قيم معالم هذا المتغير تشير إلى :

$$\sigma = 9.24 \;;\; \mu = 13.034$$

وبتوظيف القيمتين أعلاه وصيغة العلاقة اعلاه وهي :

$n = NZ^2\sigma^2/\left[Nd^2 + Z^2\sigma^2\right]$، يكون لدينا :

- بالنسبة للمنشآت التي يعمل بها أقل من 5 مشتغلين: لدينا $N= 4266$ وبفرض أن مقدار الفرق المسموح به بين μ و $\overline{X}$ هو 0.062 أي أن $d^2=0.65324$ وعند درجة ثقة 0.95 فإن $n=450$ أو ما يعادل نسبة 10.55٪ من N.

- المنشآت التي يعمل فيها 5-9 مشتغلين: لدينا $N=1224$ وبفرض فرق بين μ و $\overline{X}$ مقداره 0.08 أي $d^2=1.0876$ ودرجة ثقة 0.95 فإن $n=24$ أو ما نسبته 19.77٪ من حجم المجتمع N.

(3) اسلوب المعاينة

- اسلوب الاختيار المتناسب: أن يكون عدد الوحدات التي يتم سحبها متناسباً وحجم الطبقة: $W_i = N_i/N = n_i/n$

- ♦ أسلوب نايمن (المثالي): ويعتمد فكرة الأخذ بنظر الاعتبار مقدار التباين في كل طبقة أي : $n_i = n\left(N_i\sigma_i / \sum N_i\sigma_i\right)$

اي، لدينا :

- عينة طبقية عشوائية متعددة المراحل، ومن غير الممكن توفير المعالم المطلوبة لكافة الأنشطة والمستويات (الحد الرابع) في حالة استخدام طريقة نايمن.
- إن عملية تقسيم المجتمع إلى طبقات ولمراحل متعددة من شأنه تحقيق التجانس في كل طبقة من كل مرحلة.
- إن أهمية إيجاد تمثيل لكل نشاط ومستوى قد لا يتحقق في الطريقة المثالية لأن الزيادة في حجم أي طبقة أو نشاط من شأنه التأثير على حجم الطبقات والأنشطة الأخرى.

ومن أعلاه نجد من المناسب استخدام الطريقة الأولى (أسلوب الاختيار المتناسب) مع المسح موضوع الدراسة.

(4) اختيار وحدات المعاينة

لأجل تحقيق العشوائية بجانب التقليل ما أمكن من انتشار الوحدات المشمولة بالمسح، فقد تم الاستعانة بطريقة العينة العشوائية النظامية في عملية تحديد الوحدات التي يتم شمولها، والتي تتمثل بالخطوات التالية:

- ♦ إيجاد طول الدورة العشوائية L.
- ♦ تحديد البداية العشوائية ضمن طول الدورة.

<u>على سبيل المثال :</u> نشاط المطاعم والفنادق n=74 , N= 549 فيكون لدينا L=549/74 = 7. فنختار عشوائياً أحد الأرقام التي تقع بين 7-1، ولنفترض ظهور الرقم 5 فسيكون أرقام المنشآت التي يتم سحبها هي 5 ,12 ,19 ,26 ولغاية الحصول على 74 منشأة.

3-10-2 حالة دراسية رقم C_{3-3} :

في تحديد حجم عينة مسحوبة من مجتمع مجهول المعالم

في دراسة تستهدف الوقوف على اراء ومدى رضا (satisfaction) شريحة من المعنيين في العمل البحثي العاملين في جامعات دول عربية، وتشخيص العوامل والظروف والمعوقات التي تؤثر على النشاط البحثي وتطويره، وبسبب مجهولية معالم المجتمع لدى الباحث ليتسنى تحديد حجم العينة المناسب مسبقا، فقد تم جمع معطيات من عينة حجمها 74 استبانة من تدريسيين يعملون في جامعات عراقية واردنية واماراتية ويمنية موزعين كالاتي : العراق 50.5٪، الاردن 25.7٪، الامارات 13.5٪، اليمن 10.3٪ تدريسيا. بينهم (32.4٪) من الاناث و (67.6٪) ذكور. وتم تقسيم مستوى الرضا الى 5 درجات، اعطيت بموجبها القيمة 5 لمستوى الرضا التام نزولا عند القيمة 1 لمستوى الرضا الضعيف.

ولاجل التاكد من ان حجم العينة التي ستخضع للتحليل هي مستوفية لمعايير الدقة المطلوبة، فقد تم اعتماد متغير معدل دخل العائلة الشهري كمتغير رئيسي لتاثيره الواضح في سلوك وقرارات اي الشخص، في قياس حدود الثقة (Confidence limits) عند درجة تقة 95 ٪ والتي صيغتها العامة هي :

$$\bar{x} - t_{(1-\alpha/2)},\nu\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{(1-\alpha/2)},\nu\frac{s}{\sqrt{n}}$$

حيث ان :

القيمة الجدولية (الملحق 3-3) بدرجة ثقة 95٪ وعند درجات حرية $\nu = 73$

هي : $t_{(1-\alpha/2)}, v=1.991$

s تشير الى الانحراف المعياري للعينة

μ متوسط المجتمع

$\bar{x}$ متوسط العينة (معدل الدخل الشهري للعائلة)

وبالرجوع الى المعطيات لدينا :

$\bar{x}=$ 1290 دولار شهريا

$s=$ 687.73

$n=$ 74

والتعويض بالصيغة اعلاه نحصل على: 1290 ∓ 156.693

وبذلك نجد بان متوسط المجتمع μ يقع بين القيمتين 1133.31 دولار كحد ادنى و 1446.7 دولار كحد اعلى عند درجة ثقة مقدارها 95 ٪ .

3- 11 العينات غير العشوائية Non - Random Sample

3- 11- 1 مفهوم العينات واستخداماتها

وفيها يتم اخيار العينة من المجتمع حسب وجهة نظر الباحث من دون الاعتماد على اسس علمية تتيح القدرة على استخدام نتائج هذه العينات لغرض تعميمها على المجتمع الكلي ولا التحقق من دقة نتائجها. وتستخدم هذه العينات في الغالب لاغراض استطلاعات الراي، حيث لايترتب على نتائجها اتخاذ قرارات تشكل خطورة على موضوع الدراسة.

3- 11- 2 انواع العينات الغير عشوائية (غير الاحتمالية)

(1) العينات التحكمية (المتعمدة) Judgement Samples

ويتم اختيار وحداتها على وفق رغبة ووجهة نظر الباحث .

(2) العينات الحصصية Quata Samples

وهي العينات التي يتم اختيار وحداتها على وفق تركيبة المجتمع ،فاذا كان المجتمع يتكون مثلا من ثلاث فئات حسب دخل الاسر، ولتكن: اسر ذات دخل واطئ واسر ذات دخل متوسط، وثالثة عالية الدخل، وكانت نسب كل من هذه الفئات في المجتمع هي: 40 ٪ و 50 ٪ و 10 ٪ على التوالي، فان اختيار عينة تتكون من 100 اسرة سيتم على وفق هذه النسب لتشتمل على 40 اسرة من الفئة الاولى و 50 اسرة من الثانية و20 اسرة من الفئة الثالثة، الا ان عملية هذه الاسر تتم بصورة كيفية من الاعتماد على الاسلوب العشوائي .

الفصل الرابع
تدقيق أخطاء نتائج المسح
وطرق معالجة وتعويض المعطيات المفقودة
Survey Errors Varifying and Missing Data Imputation

1-4 تدقيق الاستبانات الاحصائية

قبل اخضاع معطيات المسح او الاستقصاء الاحصائي لمرحلة التفريغ والتبويب، يستلزم الامر المرور بسلسلة اجراءات تدقيقية لتشخيص الاخطاء الحاصلة واسلوب معالجتها للاطمئنان الى ان الاسئلة المدونة في الاستبانة قد اجيب عليها، فضلا عن احتمال الكشف عما ان كان المبحوثين قد وقعوا بأخطاء اثناء اجابتهم نتيجة الالتباس او الغموض الذي قد يكتنف بعض الاسئلة، وقد يكون بعض هذه الاخطاء حسابية وبعضها موضوعية كأن ياتي معدل انفاق الاسرة اكثر من دخلها او ان ياتي عمر الابن اكبر من عمر الاب وما شابه. ويمكن اجمال اهم اساليب التدقيق في هذه المرحلة بما ياتي :

4-1-1 اجراء مطابقة بين عدد الاستبانات المتوفرة مع اطار المسح، للتاكد من شمول كافة وحدات العينة المقرر شمولها، والبحث عن اسبابها ومسوغاتها ان وجدت .

4-1-2 مطابقة المعطيات مع المفاهيم المعتمدة في المسح، كان تكون قيمة سلعة ما هي لوحدة واحدة والمطلوب هو القيمة لدرزن مثلا او العكس صحيح

4-1-3 اجراء تدقيق حسابي كتقسيم قيمة السلعة على كميتها لتدقيق سعرها، او القيام بجمع الاجزاء للتاكد من صحة المجموع .

4-1-4 اجراء تدقيق منطقي كوضع حدود دنيا وعليا لاسعار بعض المواد الواردة في الاستبانة، او نسب بعض التوزيعات مثل تناسب الشهادة الدراسية الحاصل عليها الشخص مع عمره او تناسب عمر الابن مع عمر ابيه مثلا، وهو ما يعرف في عمل الحاسوب بالتدقيق النوعي Validation.

4-2 تدقيق نتائج اخطاء المسح (الاستقصاء)

4- 2- 1 اخطاء المعاينة (الاخطاء العشوائية) Sampling errors

وهي الاخطاء الناتجة عن الفروق العشوائية بين القيم المختارة بالعينة وباقي القيم التي شاءت الصدفة عدم اختيارها لتكون ضمن العينة، وهي تشير الى مستوى دقة نتائج العينة او المدى الذي تختلف فيه هذه النتائج مع تلك التي تستخرج من المجتمع الاحصائي الكلي، وعلى افتراض ان منهجية القياس ثابتة. اي ان هذه الاخطاء هي نتيجة لاعتماد جزء من المجتمع الكلي. ويمكن تقليل هذا النوع من الاخطاء وحساب مقدار تاثيره اذا اختيرت العينة العشوائية المناسبة لحالة المجتمع المدروس. كما يؤدي زيادة حجم العينة الى تناقص هذا النوع من الاخطاء. فمثلا لو كان لدينا مجتمع مكون من القيم الاتية : 8,9 , 12 , 25 , 13 , 28 , 10 واعتمدنا على عينة بحجم n=2 فقد نحصل على العينة بقيم 25 , 28 فيكون تقدير متوسط المجتمع هو $\bar{x} = 26.5$ في حين متوسط المجتمع الحقيقي هو $\mu = 15$ وهي قيمة بعيدة عما حصلنا عليه من العينة، وهذا راجع الى اننا لم نقم باختيار العينة العشوائية الممثلة للمجتمع المدروس. وبذلك فان حجم الخطأ يتوقف على اسلوب تصميم العينة المناسبة (كان تكون طبقية او عنقودية ... الخ) وعلى حجم العينة .

4- 2- 2 اخطاء غير المعاينة (اخطاء التحيز)
Non-Sampling errors

رغم اتباع الطرق العشوائية الصحيحة في اختيار العينة، الا انه قد يقع نوع اخر من الاخطاء سواء اكان مقصود او غير مقصود، ويسمى باخطاء التحيز، ويحصل ذلك نتيجة الاستعاضة عن بعض الوحدات المحددة في العينة لعدم استجابتها الكلية او الجزئية عند ملئ الاستبانة او خلال التبويب والتحليل. وهذا النوع من الاخطاء اكثر خطورة من الاخطاء العشوائية لاننا لانستطيع حساب مقدار تاثيره وهو يزداد بزيادة حجم العينة، ويمكن تلخيص مصادره بما ياتي :

(1) اخذ عينة من مصدر خاطئ اي اختيار عينة بالاعتماد على اطار احصائي غير صالح، كان يستعان بدليل الهاتف للحصول على عينة من تجار بيع المفرد (التجزئة) في حين قد لايتضمن الدليل جميع تجار التجزئه، اما لعدم توفر هواتف لدى قسم منهم او لعدم ورود اسمائهم في الدليل رغم امتلاكهم للهاتف .

(2) اسقاط بعض وحدات العينة عند جمع المعطيات، مثال ذلك اخذ عينة من قبل طبيب لدراسة الحالة الصحية لعمال مصنع ما والتغاضي عن الغائبين رغم ان المرض هو احد اسباب الغياب عن العمل .

(3) اخطاء ناتجة عن عدم اختيار عينة ممثلة لمجتمع الدراسة، كان نكتفي عند دراسة نمط استهلاك الاسرة، الاعتماد على الاسر في الحضر من دون شمول الريف، مع ان النمط الاستهلاكي لكل منهم مختلف .

(4) التحيز في حساب حقائق المجتمع من نتائج العينة، فاذا كنا بصدد تقدير الدخل الشهري مثلا لمنطقة سكنية تضم طبقتين من الاسر، وكانت الطبقة الاولى مكونة من 100 اسرة، ومتوسط دخلها 250 دولار، والثانية تضم 200 اسرة متوسط دخلها الشهري 180 دولار، فان التقدير الخاطئ يظهر من ايجاد المتوسط الشهري للحي من خلال جمع 250 مع 180 وقسمتها على 2 فنحصل على 215 وفي ذلك نكون قد اهملنا الاهمية النسبية لعدد الاسر لكل من الطبقتين، الذي يتطلب ترجيح متوسط دخل كل طبقة بعددها ليصبح متوسط الدخل الشهري للمنطقة هو :

$\bar{x} = (180)\ (200) + (250)\ (100)\ /\ 200{+}100 = 203.3$ دولار

(5) أخطاء القياس, وهي أخطاء ناتجة اما عن اجراءات التعامل مع المعطيات بسبب الغموض او سوء الفهم للسؤال او بسبب حب الظهور والمبالغة، او بسبب الوقوع في الخطأ عند تدوين الاجابة بصورة غير مقصودة، ومنها ما يقع خلال مرحلة الترميز او عند التفريغ الى الجداول الاحصائية.

(6) أخطاء ناتجة عن عدم الاستجابة، وهي اخطاء نظامية تؤثر في اية عينة، وتظهر هذه الاخطاء على المعطيات سواء جمعت باسلوب العينة او المسح الشامل وتكون خطورتها اكبر على الحالة الاخيرة لانه يصعب حساب مقدار تاثيرها، وتزداد بزيادة حجم العينة. وتقع اخطاء عدم الاستجابة اما بسبب عدم اجابة بعض المبحوثين على بعض اسئلة الاستبانة لعدم معرفتهم لاسئلة محددة او رفضهم الاجابة كليا. وتعود عدم الاجابة الى عدة عوامل يمكن السيطرة على بعضها بصورة معقولة كما هو الحال في حسن اختيار الباحثين وتدريبهم، وعلى اسلوب صياغة اسئلة الاستبانة، وعلى جودة

الاطار الاحصائي، وعلى التوقيت الزمني المناسب لمقابلة المبحوثين. لكن ثمة اخطاء اخرى ليس امامنا سوى حدود ضيقة للسيطرة عليها كأدعاء المبحوث عدم المعرفة .

وتزداد خطورة المشكلة اذا كانت القيم المفقودة نتيجة عدم الاستجابة تختلف عن وحدات المشاهدة التي تمت استجابتها. فعلى فرض كانت نسبة الاستجابة من مجتمع هي W_i , ونسبة عدم الاستجابة في المجتمع هي W_o فان متوسطي كلا المجتمعين بالنسبة للمتغير المطلوب قياسه هما : Y_i و Y_o عندها فان عينة المستجيبين التي حجمها n_i سيكون متوسط مربعات خطأ متوسطها y_i هو :

$$E(\bar{y}_i - \bar{Y})^2 = E(\bar{y}_i - [W_i \bar{y}_i + W_o \bar{y}_o]^2$$

وبما ان : $W_o + W_i = 1$

فان :

$$= E[\bar{y}_i - W_i \bar{Y}_i - W_o \bar{Y}_o]^2$$

$$= E[\bar{y}_i - \bar{Y}_i(W_i + W_o) + W_i \bar{Y}_i - W_o \bar{Y}_o]^2$$

$$= E[\bar{y}_i - \bar{Y}_i + W_o(\bar{Y}_i - \bar{Y}_o)]^2$$

$$= E(\bar{y}_i - \bar{Y}_i)^2 + W_o^2(\bar{Y}_i - \bar{Y}_o)^2$$

حيث ان :

$$E(\bar{y}_i - \bar{Y}_i)(\bar{Y}_i - \bar{Y}_o) = 0$$

فان :

$$E(\bar{y} - \bar{Y})^2 = \frac{s^2}{n_i} + W_o^2(\bar{Y}_i - \bar{Y}_o)^2 + B^2$$

حيث ان B تمثل هنا مقدار التحيز .

ولمعالجة مثل هذه الحالة فاما ان نخفض النسبة W_o وذلك من خلال محاولة الاتصال المباشر بالمشاهدات غير المستجيبة او باخذ نسبة المعطيات المؤدية الى التحيز بنظر الاعتبار. فلو افترضنا ان اجابات الذكور تختلف كثيرا عن طبيعة اجابات الاناث على اسئلة الاستبانة وان العينة مكونة من 49 من الذكور و 51 من الاناث، وان استجابة الذكور كانت 41 (84٪) واستجابة الاناث كانت 49 (96٪) عندها يكون حساب متوسط العينة كالاتي :

$$\bar{x} = 0.49\ \bar{x}_m + 0.51\ \bar{x}_j$$

حيث ان : $\bar{x}_m$ هو متوسط الذكور في العينة و $\bar{x}_j$ يشير الى متوسط الاناث في العينة. ويساعد مثل هذا التقدير في تخفيض مقدار التحيز.

3-4 طرق معالجة وتعويض المعطيات المفقودة

تطرقنا في اعلاه الى امكانية الرجوع الى المبحوثين غير المستجيبين كليا او جزئيا كمحاولة لمعالجة النقص في المعطيات، الا ان ذلك يواجه صعوبات العثور على هؤلاء المبحوثين احيانا او انه تكلفته تكون عالية نسبيا اضافة الى احتمال رفضه الاستجابة مجددا الامر الذي يبقي موضوع نقصان حجم العينة والاضطرار الى مراعاة ذلك عند التحليل، اما الاسلوب الاخر في حالة تعذر هذا الاسلوب في المعالجة فهو الاعتماد على طرق التعويض لتكملة المعطيات، ويمكن اجمال اهم هذه الطرق بما ياتي :

4- 3- 1 طريقة التعويض الاستنتاجي (الاستنباطي) Deductive Imputation

وهي الطريقة التي يتم فيها اشتقاق البيان المفقود من خلال نمط الاجابات على الاسئلة الاخرى في الاستبانة في حالة عدم الاستجابة الجزئية، فاشتقاق معدل الاجر الشهري للشخص يمكن ان يتم من خلال معطيات تحصيله الدراسي ومدة خدمته الوظيفية وطبيعة عمله مثلا وهكذا .

4- 3- 2 طريقة التعويض باستخدام المعدل العام

Overall Mean Imputation

وذلك باعتماد المتوسط العام لقيمة المتغير وفقا للاستبانات المتوفر فيها معطيات من قبل المبحوثين الاخرين، فاذا كان حجم العينة الاصلية للمسح مثلا $n = 300$ وان عدد قيم المعطيات التي وفرها المسح فعليا عن متغير الدخل x_i هي $m = 291$، فيعطى لكل من القيم ل 9 المفقودة قيمة المتوسط العام للقيم المتوفرة ومقداره :

$$\bar{x}_m = \frac{\sum_{i=1}^{291} x_i}{m}$$

4- 3- 3 التعويض باستخدام معدل فئة الرقم المفقود

Class Mean Imputation

وفيه يتم تصنيف العينة الى فئات على وفق متغيرات المسح ومن ثم التعويض عن القيمة المفقودة للمتغير المعني بمتوسط الفئة التي ينتمي اليها. فاذا افترضنا ان المتغير x_i الذي يعود لعينة حجمها $n = 300$ يضم ثلاث فئات وان عدد قيم كل من هذه الفئات هي: $n_1=146, n_2=94, n_3=60$ وان القيم ل 9 المفقودة، 5 منها تقع ضمن n_1 و 4 قيم تقع ضمن الفئة n_3، فان عملية التعويض لكل من القيم ال 5 تكون بالقيمة :

$$\bar{x}_{m1} = \frac{\sum_{i=1}^{141} x_{i1}}{m_1}$$

والتعويض عن كل من القيم الابعة المفقودة في الفئة الثالثة بالقيمة :

$$\bar{x}_{m3} = \frac{\sum_{i=1}^{56} x_{3i}}{m_3}$$

4- 3- 4 التعويض العشوائي العام Random Overall Imputation

وبموجبها يتم اختيار احدى الوحدات المتوفرة بطريقة عشوائية من بين وحدات العينة المستجيبة وعلى غرار طريقة الاختيار في حالة العينة العشوائية البسيطة، لتكون القيمة المختارة تعويضا عن القيمة المفقودة .

4- 3- 5 التعويض العشوائي ضمن الفئة
Random Imputation Within Class

وذلك بتقسيم العينة الى فئات متجانسة لحصر القيمة المفقودة ضمن فئة محددة ومن ثم اختيار قيمة لاحد المستجيبين عشوائيا من تلك الفئة والتعويض بها عن القيمة المفقودة .

4- 3- 6 طريقة المسافة التوفيقية
Distance Function Matching Imputation

ويتم تعوض القيمة المفقودة بموجب هذه الطريقة بقيمة اقرب وحدة مبحوثة مستجيبة وفرها المسح .

4- 3- 7 التعويض باستخدام طريقة الانحدار
Imputation Using Regression Analysis

وكما هو معلوم يمكن ايضا استخدام الانحدار لغرض التقدير والتنبؤ بقيمة متغير ما بدلالة متغير اخر، حيث تشمل المعادلة المتغير المطلوب تقديره او التنبؤ بقيمته ويطلق عليه بالمتغير التابع y دلالة المتغير المستقل x_i وهوالمتغير الذي يقوم بتحديد حصيلة القيمة المقدرة وذلك بموجب الصيغة التالية :

$$y_i = a + b\, x_i$$

حيث ان :

y_i المتغير التابع

a المعامل الثابت

b معامل الانحدار

x_i المتغير المستقل

وبضرب المعادلة اعلاه ب n يكون لدينا :

$$\sum y_i = na + b \sum x_i$$

وبالقسمة على n

$$\frac{\Sigma y_i}{n} = a + b \frac{\Sigma x_i}{n}$$

نحصل على :

$$\bar{y} = a + b\bar{x}$$

$$a = \bar{y} - b\bar{x}$$

حيث ان: $\bar{y}$ و $\bar{x}$ هما متوسطي العينة للمتغير التابع والمستقل على التوالي .

اما قيمة المعامل b يتم تقديرها باستخدام الصيغة التالية :

$$b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

وبتعويض قيمة a يمكن تقدير y على النحو الاتي :

$$\hat{y}_i = \bar{y} - b\bar{x} + bx$$

$$= \bar{y} + b(x - \bar{x})$$

4-4 استخدام النسب في تقدير مجموع N ومتوسط المجتمع μ
Using Ratios For Estimation N & μ

وهي الطريقة التي يمكن بواسطتها اعتماد النسب للحصول على تقديرات مجموع ومتوسط المجتمع من العينات. واختلافها عن تقدير نسبة خاصية ما التي تكون النسبة P فيها تمثل عدد الوحدات التي تحمل الصفة مقسومة على العدد الكلي للعينة، هو ان النسبة في الطريقة موضوعة البحث يكون مقامها قيمة متغير من عينة اخرى بشرط ان يكون هناك ارتباط بين قيمة البسط ولتكن x وقيمة المقام ولتكن y، كتقدير مجموع الانتاج الى مجموع المساحة المزروعة واستخدام النسبة في تقدير الانتاج الكلي بمعلومية المساحة الكلية .

فلو افترضنا ان لدينا عينة لتقدير مجموع المجتمع بالنسبة للمتغير y_i وكانت هناك معطيات متوافرة لنتائج سابقة عن المتغير نفسه ولنرمز له بـ x_i فان النسبة ratio ستكون عبارة عن :

$$\text{Ratio } (r) = \frac{\Sigma y_i}{\Sigma x_i} = \frac{\bar{y}}{\bar{x}}$$

حيث ان Σy_i هو مجموع قيم المتغير y_i للعينة الحاضرة و Σx_i هو مجموع قيم المتغير x_i للعينة السابقة. ويصبح بالامكان استخدام هذه النسبة في تقدير المجموع الكلي السابق، فاذا رمزنا لمجموع المجتمع الحاضر Y ولمجموع المجتمع السابق X فان ضرب النسبة ratio بمجموع المجتمع السابق المعلوم X نحصل على مجموع المجتمع الحاضر التقديري، أي :

$$\hat{Y} = r\left(\sum_{i=1}^{N} X_i\right) = \left(\frac{\bar{Y}}{\bar{X}}\right)\left(\sum_{i=1}^{N} X_i\right)$$

ومن الطبيعي انه كلما كانت مستوى التغيرات الحاصلة لكلا الحالتين الحاضرة والسابقة متقاربا او متشابها ازدادت دقة نتائج تقديرات معالم المجتمع.

اما الخطأ المعياري التقديري للتقدير $\hat{Y}$ فان نتائج العينة العشوائية البسيطة التي يكون حجمها n كبيرا نسبيا ستكون مقاربة لحصيلة الصيغة التالية في حساب الخطأ المعياري لمجموع المجتمع :

$$S_{\hat{Y}} = N \sqrt{\frac{\Sigma(y_i - rx_i)^2}{n(n-1)}} \sqrt{\frac{N-n}{N}}$$

حيث ان r ترمز الى ratio

اما الخطأ المعياري للنسبة r في حالة العينات الكبيرة سيكون على النحو الاتي

$$S_r = \frac{1}{\bar{x}} \sqrt{\frac{\Sigma(y_i - rx_i)^2}{n(n-1)}} \sqrt{\frac{N-n}{N}}$$

مثال (1.4): المعطيات التالية تبين عدد السكان (بالمليون) لسنتي 1996 و 2006 لست مدن N= 6 والمطلوب حساب تقدير مجموع المجتمع لسكان 2006 باستخدام عينة عشوائية حجمها n = 2 بالاعتماد على نسبة السكان لسنة 1996 مع ايجاد تقدير الخطأ المعياري $S_{\hat{Y}}$ وكذلك تقدير الخطأ المعياري للنسبة S_r .

Total	6	5	4	3	2	1	year
ΣY=18.77	1.23	1.51	1.95	7.9	2.82	3.36	2006 (Y)
ΣX=18.42	0.94	1.67	2.0	7.7 8	2.48	3.35	1996 (X)

الحل لـ (1.4) :

لدينا $n = 2$, $N = 6$ فيكون مجموع العينات هو :

$$\binom{N}{n} = 15$$

وهي كما مبين في الجدول التالي :

$(y_i - r\,x_i)^2$	y_i (2006)	x_i (1996)	sample
0.0022	6.19	6.03	1,2
0.0857	11.26	11.33	1,3
0.116	5.31	5.55	1,4
0.192	4.87	5.22	1,5
0.005	4.59	4.49	1,6
0.081	10.72	10.26	2,3
0.046	4.77	4.48	2,4
0.012	4.33	4.15	2,5
0.330	4.05	3.42	2,6
0.009	9.85	9.78	3,4
0.04	9.41	9.45	3,5
0.68	9.13	8.72	3,6
0.11	3.40	3.67	4,5
0.036	3.18	2.94	4,6
0.007	2.74	2.61	5,6

المجموع : $\Sigma x_i = 92.1$ $\qquad \Sigma y_i = 93.7$

المتوسط : $\bar{x} = 6.14$ $\qquad \bar{y} = 6.247$

$$r = \frac{\Sigma y_i}{\Sigma x_i} = \frac{\bar{y}}{\bar{x}} = \frac{6.247}{6.14} = 1.0174$$

$$\hat{Y} = \text{ratio } (\sum_{i=1}^{N} x_i) = (\frac{\bar{y}}{\bar{x}})(\sum_{i=1}^{N} x_i)$$

$$= (\frac{6.247}{6.14})(18.42) = 18.741$$

$$S_{\hat{Y}} = N \sqrt{\frac{\Sigma(y_i - rx_i)^2}{n(n-1)}} \sqrt{\frac{N-n}{N}}$$

$$= 6(0.753)(0.89) = 4.021 \sqrt{\frac{1.1358}{2}} \sqrt{\frac{4}{6}} = 6$$

$$S_r = \frac{1}{\bar{x}} \sqrt{\frac{\Sigma(y_i - rx_i)^2}{n(n-1)}} \sqrt{\frac{N-n}{N}}$$

$$= \frac{1}{6.14}(0.753)(0.89) = 0.109$$

4- 5 حالة دراسية رقم C_{4-1}

في استخدام الانحدار في التقدير والتعويض

في مسح بالعينة استهدف جمع معطيات لاغراض تخطيط النقل شمل $n = 842$ مسافرا. وبعد القيام بمرحلة التدقيق، اتضح ان هناك قيم مفقودة كان اغلبها بسبب عدم الاستجابة، ومن بين تلك المعطيات المفقودة ما يخص متغير طول فترة الانتظار (waiting time)، بلغت عدد هذه القيم المفقودة في 57 استبانة غير مكتملة الاجابة عن هذا المتغير. وبغية توفير هذه القيم المفقوده، كان لابد من اللجوء الى التقدير لاستحالة الرجوع الى وحدات

العينة غير المستجيبة جزئيا، وذلك بسبب كون المجتمع الذي سحبت منه العينة وهو مجتمع المسافرين يعتبرمجتمعا متحركا يصعب معرفة عنوانه او ان الوصول الى الوحدات غير المستجيبة ان توفرت بعض عناوينهم ستكون مكلفة جدا .

وكان لابد من تكملة جدول المدخلات ليكون جاهزا لاخضاعه لعملية التحليل، لذلك فقد تم الاستعانة بالاستبانات الكاملة الاجابة التي وفرها المسح بما فيها معطيات المتغير المذكور لتطوير نموذج انحدار لغرض استخدامه في تعويض لـ 57 قيمة مفقودة، واعتبار المتغير المفقود دالة (y) في قيم المتغيرات المستقلة (x_i). وقد شملت عملية التحليل لبناء نموذج الانحدار كل من المتغيرات المستقلة التالية :

(1) العمر (عمر المسافر بالسنين)

(2) الجنس (القيمة 0 للاناث، 1 للذكور)

(3) المهنة (ويتكون من ثلاث فئات : 0، 1، 2)

(4) معدل الدخل الشهري للمسافر (بالدولار)

(5) هدف السفر (رحلة عمل 0، رحلة غير عمل 1)

(6) يوم السفر (ايام الاسبوع : 1، 2، ....، 7)

(7) وقت السفر (قبل الظهر 0، بعد الظهر 1)

(8) اجور النقل (بالدينار او اجزائه)

(9) طول زمن الرحلة (بالساعة او اجزائها)

(10) طول مسافة الطريق، خط السير (بالكيلومتر)

(11) نوع واسطة النقل المستخدمة (صالون 0، باص متوسط الحجم 1، حافلة كبيرة 2)

وباستخدام معادلة الانحدار التقديرية المبينة صيغتها في ادناه:

$$\hat{y} = a_o + a_1x_1 + a_2x_2 + \ldots\ldots + a_kx_k + e_i$$

حيث ان :

$\hat{y}$ المتغير التابع المستهدف تقدير قيمه المفقودة (طول فترة الانتظار) وتقع قيمه بين 1 و 9 والاخيرة تمثل قيمة اعلى اهمية وفقا لراي وحدة العينة (المسافر).

a_o المعامل الثابت

a_i معاملات الانحدار

x_i المتغيرات المستقلة

e_i المتغير العشوائي (الاخطاء) اللازم اختبار فرضياتها (الموضحة تفاصيلها في الفصل السادس). وان k تدل على عدد المتغيرات المستقلة $i = 1, 2\ldots, k$

ملاحظة :

ان الخطوات التحليلية التالية مبينة تفاصيلها في الفصل السادس، ونشير اليها هنا باختصار تماشيا مع تسلسل المواضيع، لذا يفضل متابعة هذه الحالة الدراسية بعد الرجوع الى الفصل المذكور والمتعلق بتحليل الانحدار المتعدد .

(1) بناء النموذج: بعد التخلص من العلاقات المتداخلة بين المتغيرات المستقلة (Multicollinearity) من خلال فحص مصفوفة الارتباط، تلتها

مرحلة فحص معايير المعنوية على مخرجات النموذج والتي تشمل كل من: معامل الارتباط المتعدد R، معامل التحديد R^2، الاخطاء المعيارية .S.E و F و مستوى المعنويه α. كانت الحصيلة هي تطوير نموذج الانحدار التالي :

$$\hat{y} = -1.65 + 74.89\, x_1 + 0.249\, x_2$$

مربعات الخطأ	(0.02)	(2.02)	(3.46)
مستوى المعنوية	(0.000)	(0.000)	(0.000)
معامل الارتباط المتعدد	$R = 0.83$		
معامل التحديد	$R^2 = 0.69$		
مربعات الخطأ للنموذج	S.E. = 25.99		
معامل اختبار النموذج	F = 776.85 Sig. at 0.000		

(2) فحص معايير دقة النموذج وجودته : وكما يتضح فان النموذج يتضمن متغيرين عالية المعنوية من مجموع 11 متغيرا تم اخضاعها لعملية التحليل، وهما كل من: متغير واسطة النقل المستخدمة (x_1) ومعدل الدخل الشهري لاسرة المسافر (x_2)، وكلاهما جاء باشارة صحيحة، ومقبول منطقيا لكون متغير الدخل يمثل احد اقوى المتغيرات المؤثرة على اي متخذ قرار من جهة وكونة ممثلا لمعظم المتغيرات التي لم يتضمنها النموذج، اما متغير واسطة النقل فان اختيار نوع واسطة السفر يرتبط مباشرة بطول فترة الانتظار، فاختيار الصالون مثلا يعني الانتظار لوقت قصير بالنظر لقلة عدد الركاب المطلوب توفرهم لانطلاقها، والباص ذات الحجم المتوسط بحاجة لوقت انتظار اقل من الحافلة الكبيرة وهكذا .

(3) معايير فحص استيفاء النموذج للفرضيات جاءت ايضا عالية المعنوية من خلال فحص الاشكال البيانية المبينة في (1.4) و(2.4) و (3.4) و (4.4).

(4) اختبار قدرة النموذج المطور في بناء تقديرات القيم المفقودة :

والاختبار يتم من خلال مقارنة نتائج القيم المستخرجة بموجب النموذج $\hat{y}$ مع القيم الحقيقة للمتغير التابع y، واستخدام معيار الانحراف الطبيعي Normal deviate والذي صيغته :س

$$ND = \frac{\sum\left(e_i - \bar{e}\right)^2}{s_e}$$

حيث ان :

e_i قيم الاخطاء (الفروق بين القيم الحقيقية والقيم التقديرية)

$\bar{e}$ متوسط قيم الاخطاء

S_e الانحراف المعياري لقيم الاخطاء

نجد ان الفروق e_i التي تقع خارج ± 1.96 لاتزيد على 0.04 من حجم العينة مما يدل على قبول النتائج بدرجة ثقة مقدارها 95 ٪ . والجدول التالي رقم (1.4) يعطي مقطعا من نتائج المقارنة بين القيم الحقيقية والقيم التقديرية المستخرجة بواسطة النموذج عند مستوى معنوية $\alpha = 0.05$.

جدول رقم (1.4)

مقطع من نتائج المقارنة بين القيم الحقيقية y

والقيم التقديرية $\hat{y}$ المستخرجة بواسطة النموذج

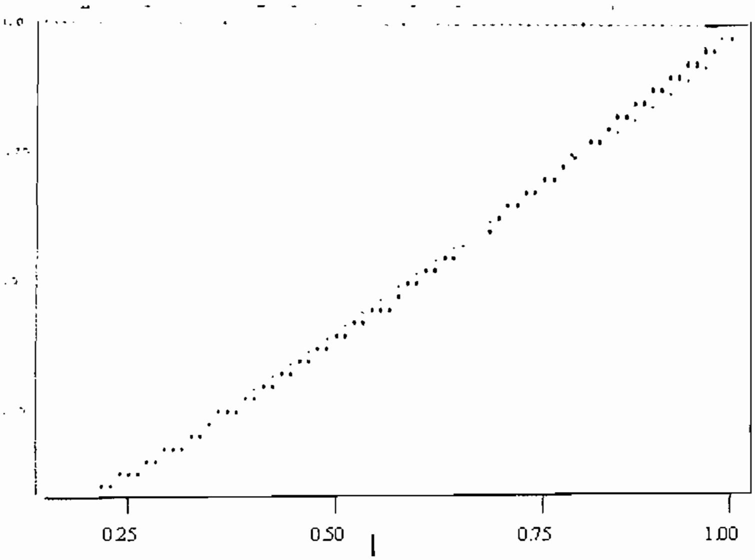
y	$\hat{y}$	$e_i = \hat{y} - y$	y	$\hat{y}$	$e_i = \hat{y} - y$
8	7.43	0.47	8	7.20	0.80
7	7.15	0.15	8	7.25	0.75
8	7.18	0.82	6	7.37	-0.63

9	8.41	0.59	9	7.88	1.12
8	8.05	-0.05	7	7.41	-0.59
7	7.75	0.75	8	7.88	0.12
8	7.30	0.30	7	7.11	-0.11
7	7.44	-0.44	8	7.90	0.10
8	7.78	0.22	8	8.23	-0.23
6	7.77	-1.77	8	7.65	0.35

ومن استعراض طرق التعويض في اعلاه نستدل على انه باستثناء طريقة الانحدار فان الطرق الاخرى تعد بسيطة وسهلة الاستخدام، وان بعضها يتصف بالاعتباطية كما هو الحال مع طريقة المسافة التوفيقية .

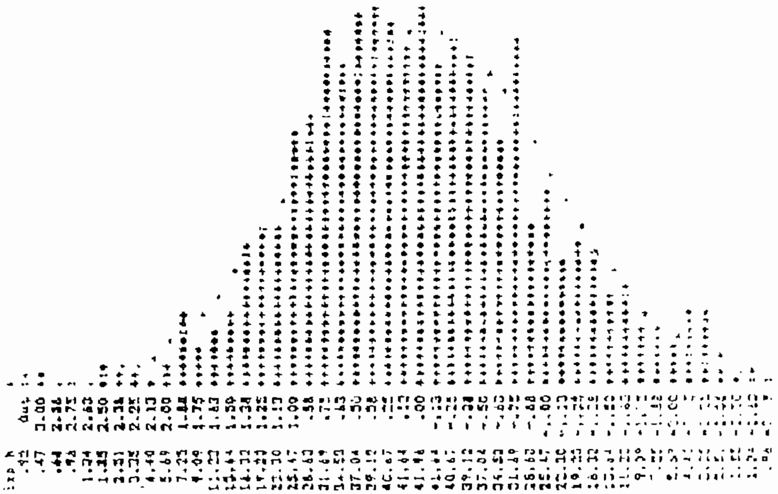
الشكل البياني رقم (1.4)

القيم الحقيقية المعيارية والنظرية للبواقي للحالة الدراسية C4-1



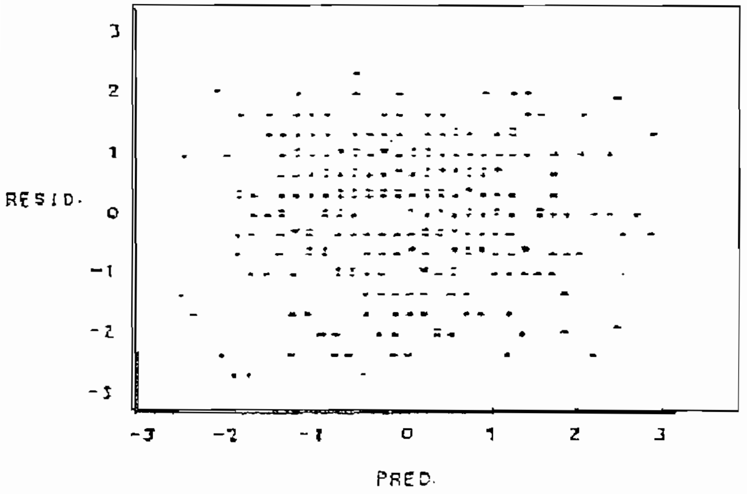
الشكل البياني رقم (2.4)

توزيع القيم المعيارية لبواقي نموذج الحالة الدراسية C4-1



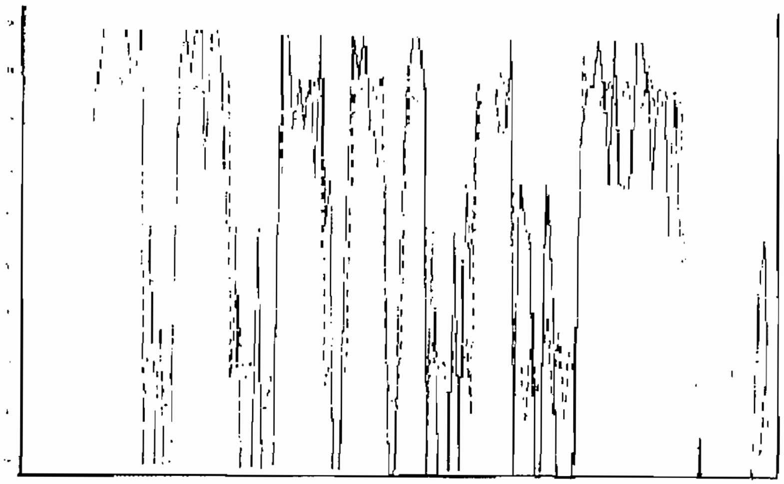
الشكل البياني رقم (3.4)

انتشار قيم البواقي لنموذج للحالة الدراسية C_{4-1}



الشكل البياني رقم (4.4)

مقارنة بين القيم الحقيقية والتقديرية المستخرجة بواسطة النموذج C4-1



الفصل الخامس

التحليل الوصفي

Descriptive Analysis

كما تم ذكره في الفصل الاول، من ان الوصف والتفسير يعني الاشارة الى سرد خصائص المعطيات مع مبررات اعتمادها في البحث، والكشف عن اتجاهها والربط بين عناصرها المختلفة. وان الادوات التي يتم استخدامها والتطرق اليها في هذا الفصل لاجل تحقيق هذه الاهداف تتمثل بالنزعة المركزية وغير المركزية، والتشتت، العرض البياني، تحليل الارتباط ، ومن ثم تحليل الانحدار، فتحليل المركبات الاساسية .

5-1 النزعة المركزية وغير المركزية والتشتت

Central, Non-Central &Dispersion

هناك خاصيتان اساسيتان لاية معطيات احصائية تساعد على اعطاء مدلول واضح لوصفها هما: الأولى هي النزعة المركزية ومقاييسها متمثلة بالمتوسطات التي بواسطتها نتمكن من تحديد موقع النقطة التي تتمحور حولها المعطيات بالاضافة للمقاييس غير المركزية المتمثلة بالعشير والربيع والمئين.

اما الثانية فهي مقاييس التشتت ويقصد بها حالة الانتشار التي تكون عليها المعطيات حول المتوسط، اي المسافات التي تبتعد فيها القيم عن المركز. بالاضافة للمقاييس الاخرى التي توفرها خواص الانحراف المعياري باعتباره احد مقاييس التشتت .

5- 1- 1 المقاييس المركزية (المتوسطات)

والمتوسط هو قيمة مفردة تمثل مجموعة من قيم المعطيات، وهناك عدة انواع من المتوسطات لكل منها طريقته الخاصة في الاحتساب، واهم هذه الانواع هي :

(1) الوسط الحسابي Arithmetic meam $\overline{X}$

ويعتبر من اهم مقاييس النزعة المركزية، وعملية حسابه غير معقدة ومفهومة ويتسم بسعة استخداماته ومن ميزاته شموله على كافة وحدات التوزيع التكراري، كما يمكن توظيفه لايجاد مجموع قيم المشاهدات عند معلومية حجم العينة، حيث ان $\Sigma x_i = n\ \bar{x}$ من خلال الصيغة التي شكلها في حالة العينة هو:

$\bar{x} = \frac{\Sigma x_i}{n}$ ، وفي حالة المعطيات البوبة $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$

حيث ان :

$\sum x_i$ هي مجموع قيم وحدات العينة

n هي عدد وحدات العينة و f_i هي التكرارات

(2) الوسيط Median , M_d

وللحصول عليه يتم اولا ترتيب المعطيات تصاعديا من الاصغر فالاكبر او تنازليا من الاكبر فالاصغر، والوسيط يصبح عبارة عن القيمة الوسطية (عندما يكون عدد المعطيات فرديا) ومتوسط القيمتين الوسطيتين (عندما تكون عدد القيم زوجية). اما في حالة التوزيع التكراري (حالة المعطيات المبوبة) فيحتاج الى اتباع الخطوات التالية :

- استخراج التوزيع التكراري الصاعد ،وهو التكرار الذي يبدأ تجميعه من الاعلى باتجاه الاسفل، اي البدأ بتكرار الفئة الاولى ومن ثم يضاف اليه تكرار الفئة الثانية ليصبح التكرار المتجمع للفئة الثانية وباضافة تكرار الفئة الثالثة يصبح التكرار المتجمع للفئة الثالثة وهكذا .

- تحديد موقع الوسيط بقسمة مجموع التكرارات على 2، اي : $\frac{\Sigma f_i}{2}$
- تحديد قيمة موقع التكرار الوسيط بين التكرارات المتجمعة .
- تحديد الفئة الوسيطة، فاذا كانت قيمة الوسيط مساوية لاي تكرار متجمع حينئذ فان فئة ذلك التكرار ستكون هي الفئة الوسيطة، اما اذا وقعت بين تكرارين متجمعين فان الفئة الاحقة لقيمة الموقع ستكون هي الفئة الوسيطة.
- نستخدم الصيغة التالية لحساب قيمة الوسيط :

$$M_d = L + \frac{\frac{\Sigma f_i}{2} - f_i}{f_2 - f_1} . H$$

حيث ان :

L : الحد الادنى لفئة الوسيط

$\frac{\Sigma f_i}{2}$: قيمة موقع الوسيط

f_1: التكرار المتجمع السابق لقيمة موقع الوسيط

f_2: التكرار المتجمع الاحق لقيمة موقع الوسيط

H: مدى (طول) الفئة

ويمتاز الوسيط بعدم تاثره بالقيم المتطرفة (الشاذة) لان موقعها سيكون في الاطراف اي بعيدة عن موقع الوسيط ذلك لاتدخل في تمثيلها للمعطيات، مع امكانية استخدامه مع الفئات المفتوحة وغير المتساوية. ويعاب عليه لاعتماده على قيمة واحدة او قيمتين و على فئة واحدة في حالة المعطيات المبوبة. كما انه قد لايعبر بصورة صحيحة عن مركز تجمع المعطيات عندما يكون عددها قليلا .

(3) المنوال Mode ، M_0

والمنوال هو القيمة الاكثر تكرارا بين مجموعة القيم، ولذلك فان قيمته قد لاتكون الوحيدة بل قد تكون هناك اكثر من قيمة منوالية واحدة، وكل منها قد يتكرر لعدة مرات. الا انه يتميز بامكانية استخدامه مع القيم الكمية والنوعية. اما في حالة المعطيات المبوبة، فان ايجاده يتطلب تحديد الفئة المنوالية التي هي الفئة التي يقابلها اكبر تكرار، ومن ثم تطبيق الصيغة التالية:

$$M_0 = L + \frac{d_1}{d_1 + d_2} . H$$

حيث ان :

L : الحد الادنى للفئة المنوالية

d_1 : تكرار الفئة المنوالية – تكرار الفئة السابقة

d_2 : تكرار الفئة المنوالية – تكرار الفئة الاحقة

H : طول الفئة .

(4) **الوسط الهندسي Geometric mean**

ويستخدم مع النسب ومعدلات النمو ومع الارقام القياسية. ويعرف من انه عبارة عن الجذر n لقيم عددها n، وصيغسة حسابه هي :

$$\bar{x}_g = \sqrt[n]{x_1 . x_2 \ldots\ldots x_n}$$

$$\log \bar{x}_g = \frac{1}{n} \Sigma \log x_i$$

اما في حالة المعطيات المبوبة فصيغة حسابه هي :

$$\bar{x}_g = \sqrt[\Sigma f]{x_1^{f1} x_2^{f2} \ldots\ldots x_n^{fn}}$$

$$\text{Log } \bar{x}_g = \frac{1}{\Sigma fi} \Sigma fi \ \log x_i$$

2-1-5 المقاييس غير المركزية

ان المتوسطات اعلاه وكما ذكرنا تستهدف تحديد المركز الذي تتمحور حوله المعطيات، وباستثناء المنوال فان جميعها تتمثل بقيمة مفردة واحدة تكون ممثلة للمعطيات التي تكون تحت الدراسة. أما في حالة المقاييس التي تحدد لنا مواقع غير مركزية كايجاد القيم عند 10٪ مثلا نستخدم معها طريقة العشريات Deciles، ولايجاد قيمة 25٪ نستخدم طريقة الربعيات Quartiles ونستخدم المئيات Percentiles اذا كنا بصدد تحديد نسبة اخرى كـأن تكون 40٪ وهكذا. وحيث ان مثل هذه التقسيمات تكون قليلة الاهمية مع المعطيات قليلة العدد (غير المبوبة) لامكانية ايجادها من دون عمليات حسابية، فنتناول في الاتي صيغها مع المعطيات المبوبة، مفترضين بان الفئات تحتوي ضمنا على اجزاء، وان المعطيات موزعة بالتساوي على مدى الفئة المعنية، وبذلك فهي تتماثل مع طريقة احتساب الوسيط.

(1) العشير Decile

فالعشير الاول هو القيمة التي تقع عند العشر $\frac{1}{10}$ الادنى من القيم، ويليه $\frac{9}{10}$ من القيم، وبنفس التعريف ينطبق على الاعشار الاخرى، وصيغة تحديد موقعه i : هي $\frac{(i)\sum f_i}{10}$

فموقع العشير الثاني مثلا هو : $\frac{(2)\sum f_i}{10}$

حيث ان i ترمز الى العشير $(i = 1,2,3,....)$، فاذا رمزنا للعشير الاول بـ D_i فان صيغة حسابه ستاخذ الشكل التالي :

$$D_i = L + \frac{\frac{(i)\sum f_i}{10} - f_i}{f_2 - f_1} . H$$

حيث ان :

L هي الحد الادنى لفئة العشير الاول

f_1 هي التكرار المتجمع الصاعد السابق لموقع العشير

f_2 هي التكرارالمتجمع الصاعد اللاحق لموقع العشير

H هي مدى الفئة

(2) الربيع Quartiles :

ان الارباع الثلاثة للتوزيع تعني تقسيم المعطيات الى 4 أجزاء كل جزء منها يشتمل على عدد متساوي من المعطيات، فاذا رمزنا للربع الاول بـ Q_1 ويقصد به المعطيات التي تقل عن Q_1 ، ونصف المعطيات تكون اقل من Q_2 (الربع الثاني)، والربع الثالث يقل عن Q_3 من المعطيات، وبذلك يفترض ان تتطابق Q_2 مع قيمة الوسيط M_d في تقسيم المساحة تحت المنحني. اما صيغ حساب موقع الربيع فهي :

$$\frac{(i)\sum f_i}{4}$$

وتحديد موقع الربيع الثالث Q_3 مثلا هو :

$$\frac{(3)\sum f_i}{4}$$

اما صيغة حساب الربيع الاول فهي :

$$Q1 = L + \frac{\frac{\sum f_i}{4} - f_1}{f_2 - f_1}.H$$

حيث ان :

f_1 التكرار المتجمع الصاعد السابق لموقع الربيع

f_2 التكرار المتجمع الصاعد اللاحق لموقع الربيع

(3) المئين Percentile :

فالمئين الاول هو القيمة الواقعة عند $\frac{1}{100}$ من قيم المعطيات والمئين 70 C_{70} مثلا هو الذي يقع عند $\frac{70}{100}$ من المعطيات وهكذا .

وان ايجاد موقع C_{30} مثلا هو :

$$\frac{30\sum f_i}{100}$$

وان صيغة حساب قيمة C_{30} هو :

$$C_{30} = L + \frac{\frac{30\sum f_i}{100} - f_1}{f_2 - f_1}.H$$

5- 1- 3 مقاييس التشتت

اما مقاييس التشتت والتي تقيس مدى ابتعاد كل قيمة من قيم اية مجموعة معطيات عن المتوسط، ومن خلالها نستطيع معرفة مستوى التجانس والاختلاف بين وحدات ظاهرة معينة او بين ظواهر متعددة عند توظيفها في بناء التقديرات في الاحصاء الاستلالي، فاهم انواعها هي :

(1) المدى Range

وهو عبارة عن الفرق بين اكبر واصغر قيمة بين المعطيات، اما في حالة المعطيات المبوبة فتكون قيمته تقديرية وذلك لمجهولية اكبر واصغر قيمة، وبذلك فان القيمة التقديرية هي عبارة عن الفرق بين الحد الادنى للفئة الدنيا والحد الاعلى للفئة العليا .

(2) الانحراف المعياري Standard deviation

ويعتبر المقياس الاكثر اهمية واستخداما للتشتت ويرمز له في حالة العينة s وفي حالة المجتمع σ وصيغة حسابه في حالة المعطيات غير المبوبة هي:

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1}}$$

اما في حالة المعطيات المبوبة فصيغة حسابه هي :

$$S = \sqrt{\frac{\Sigma f_i x_i^2 - \frac{(\Sigma f_i x_i)^2}{n}}{n-1}}$$

حيث ان : مجموع التكرارات $\Sigma f_I = n$

5- 1- 4 خواص الانحراف المعياري

Standard Deviation Properereties

(1) معامل الاختلاف (التغاير النسبي) varaition coeffecient

ان الانحراف المعياري وكذلك الحال لمقاييس التشتت الاخرى هي ذات قيم مطلقة لاتوضح مقدار التشتت في حالة اختلاف مقاييس المعطيات كالمتوسطات الحسابية، فمثلا ان قيمة التشتت لمعطيات مقاسة بالسنتيمترات هي 12 سم فان قيمتها ستكون 0.12م عند قياسها بالمتر، فتشتت مقداره 11 سم في اطوال عينة من الاشخاص يعتبر معقولا ولكن نفس المقدار من التشتت في اطوال اقدامهم يعتبر كبيرا، لان متوسط طول الشخص يبلغ عدة امثال متوسط طول قدمه وهكذا. لذلك بالامكان استخدام معامل الاختلاف ولنرمز له V كمقياس مناسب لمقارنة مقدار التشتت او الاختلاف لمجموعتين او اكثر من المعطيات في حال اختلاف اقيام الوسط الحسابي وكذلك في حال اختلاف الوحدات القياسية المستخدمة مع وحدات كل مجموعة. والصيغة التي تستخدم لهذا الغرض في حالة العينة هي :

$$V = \frac{s}{\bar{x}} . 100\%$$

(2) مقاييس التماثل والالتواء symmetry skewness measures

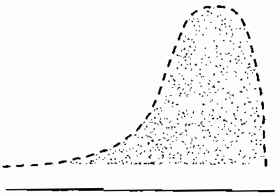
ويبحث في شكل توزيع اي مجموعة معطيات احصائية لمعرفة مدى تماثل التوزيع ومعرفة درجة الالتواء skewness واتجاهه، ويعتبر المدرج التكراري الذي منه نحصل على المنحنى افضل وسيلة للاستدلال السريع على شكل توزيع المعطيات وكما مبين في الاشكال البيانية رقم (1.5) و(2.5) و(3.5)، ومن اهم مقاييس التماثل والالتواء هو معامل بيرسن Pearsonain coefficient ونرمز له Sk، و العزم الثالث Third Moment ونرمز له M_3، وتقع قيمة معامل بيرسن بين 3± وان الاشارة تدل على اتجاه الالتواء فالقيمة السالبة تشير الى اتجاه اليسار والموجبة اتجاه اليمين، وصيغته :

$$Sk = \frac{3(\bar{x} - M_d)}{s}$$

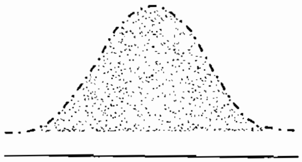
حيث ان :

$\bar{x}$ هو الوسط الحسابي و M_d الوسيط و s الانحراف المعياري

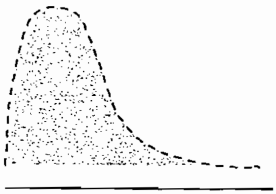
شكل بياني رقم (1.5)
يوضح توزيع معطيات متغير ملتوي باتجاه اليسار (سالب)



شكل بياني رقم (2.5)
يوضح التوزيع الطبيعي للمعطيات (متماثل)



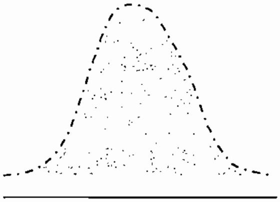
شكل بياني رقم (3.5)
يوضح توزيع معطيات متغير ملتوي باتجاه اليمين (موجب)



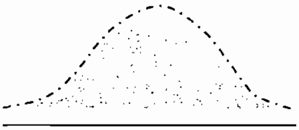
(3) مقاييس التفرطح (او التدبدب) kurtosis (or peakness)

ويقصد به درجة تدبدب قمة منحنى التوزيع، فعندما يكون شكل التوزيع ذات اطراف واسعة نسبيا وقمة ضيقة يطلق عليه بالمدبدب peakness، اما عندما تكون قمة المنحنى مسطحة فيطلق عليه بالتوزيع المفرطح Kurtosis، في حين عندما يكون التوزيع بين الحالتين نطلق عليه معتدل التفرطح Mesokurtic، والمقياس الذي يستخدم لقياس درجة التفرطح هو العزم الرابع Fourth Moment ونرمز له بـ M_4، والاشكال البيانية (4.5) و (5.5) و (6.5) تمثل نماذج من هذه التوزيعات. وللزيادة في تفاصيل خواص الانحراف المعياري وصيغها والامثلة عليها يمكن الرجوع الى كتاب المؤلف، 1997 .

شكل بياني رقم (4.5) يوضح
توزيع مدبدب (Peakness)



شكل بياني رقم (5.5) يوضح
توزيع معتدل التفرطح (Mesokurtic)



شكل بياني رقم (6.5) يوضح
توزيع مفرطح (Kurtosis)

5- 1- 5 العرض البياني Graphical Presentation

ومن الادوات المهمة والسمات الاساسية البارزة الاخرى للاحصاء الوصفي هي الاشكال البيانية حيث بواسطتها نتمكن من ايجاد عدد من المقاييس اعلاه لانها وسيلة مهمة للكشف عن اتجاه المعطيات وطبيعة توزيعها، بالاضافة الى انها تمكننا من عرض نتائج التحليل بطريقة سهلة وواضحة واكثر قبولا من الارقام، وهناك العديد من الخيارات في العرض البياني الا انه بصورة عامة يتم اختيار المناسب منها وفقا لطبيعة المعطيات ورغبة الباحث ولكن الاهم

من ذلك هو مراعاة متطلبات هدف التحليل ان كان وصفيا او تحليلا متقدما.

فعندما يكون الغرض مثلا اختبار فرضيات بناء نموذج يستلزم التاكد من توزيع البواقي Residuals لمعرفة شكل توزيعها والوقوف على شكل انتشار التباين لها وغير ذلك نلجأ الى اشكال بيانية على غرار تلك المبينة في الملاحق من (1.4) ولغاية (4.4) وغيرها من الاشكال البيانية مما سيرد في الفصل السادس، وفي مثل هذه الحالة يفضل استخدام برنامج SPSS لانه سيكون اكثر ملائمة ودقة وعادة ما يمكن الحصول عليها بصحبة بمخرجات التحليل او بتوظيف الامر الرئيسيGraph او من خلال تحديد المطلوب منها خلال اجراءات التحضير لعملية التحليل عند استخدام الامر Analysis.

اما عندما يكون الهدف من العرض هو لمتطلبات التحليل الوصفي وهو موضوعنا في هذا الفصل فان الاشكال التالية تمثل اهم الاشكال البيانية الشائعة، مع الاشارة الى امكانية الاستعانة ايضا ببرنامج Excel الذي تتوفر فيه خيارات متعددة اخرى ايضا لهذا الغرض.

(1) الاعمدة البيانية Bar charts

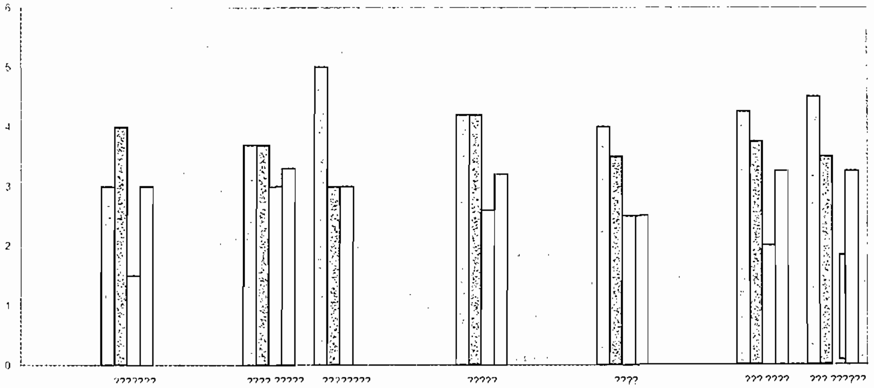
وقد تكون هذه الاعمدة على شكل اعمدة عمودية او على شكل مستطيلات افقية، وتستخدم مع التوزيعات التكرارية البسيطة او المزدوجة، وتتكون من محورين احدهما افقي يحتوي على المتغير x التي قد يمثل الزمن (سنين او اشهر ... الخ) او المكان (مدن او اقاليم او دول ... الخ) او الصفات و الخصائص (كالحالة التعليمية او صنف البضاعة او مستوى النجاح .... الخ). اما المحور العمودي y فتؤشر عليه القيم او التكرارات. وتعتبر الاعمدة البيانية هي من اكثر الانواع البيانية استخداما. وقد تكون

الاعمدة احادية في حالة هناك متغير واحد كتوزيع عدد الاشجار حسب المدن مثلا. او ذات اعمدة مزدوحة عندما تكون الظاهرة المطلوب عرضها تتكون من متغيرين كعرض الاستيردات والصادرات حسب السنين مثلا، في حين تدعى بالاعمدة المتعددة عندما تتكون الظاهرة من اكثر من متغيرين كما هو مبين في الشكل البياني رقم (7.5) الذي يضم اربعة متغيرات تخص مستوى رضا مستخدمي النقل العام عن خصائص النقل التالية: توفر الواسطة، اجور النقل، توفرها في الموقع الملائم للسكن او العمل ومدى توفر الراحة والملائمة في الواسطة موزعين حسب اسماء مناطق سكنهم او عملهم.

كما ويمكن ايضا عرض الظاهرة ذات المتغيرات المتعددة على شكل اعمدة مركبة بحيث يمثل ارتفاع العمود مجموع قيم المستويات للحالة المعنية كما مبين في الشكل البياني رقم (8.5) لنفس معطيات الشكل البياني رقم (7.5)

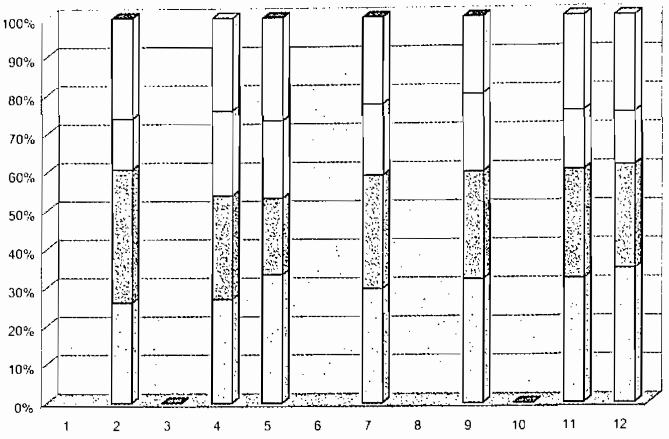
شكل بياني رقم (7.5)

يوضح نموذج الاعمدة البيانية المتعددة



شكل بياني رقم (8.5)

نموذج الاعمدة البيانية المركبة

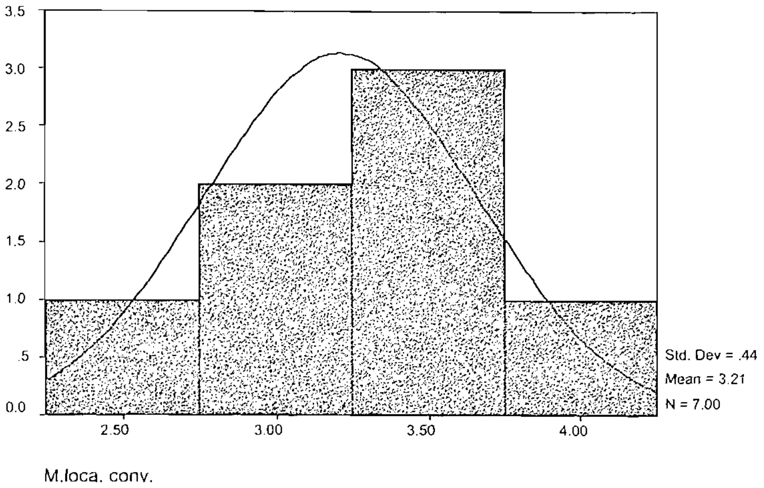


(2) مدرجات ومنحنيات التوزيعات التكرارية histogram, polygon, smoothed polygon and cummulative polygon

ومتطلبات انشاؤها هو ايضا محورين، افقي لادراج الفئات او مراكزها ومحور عمودي لتعيين تكرارات الظاهرة، وكما مع جميع الانواع الاخرى يجب تقسيم هذه المحاور الى اجزاء متساوية المديات، لان الشكل البياني يكون اقل دلالة في حالة التقسيمات او الفئات غير المتساوية. وحيث ان المساحات التي ستكون في الخارج عند رسم المنحني او المضلع من خلال التوصيل بين مراكز فئات المدرج هي مساوية للمساحات التي ستدخل تحت المنحني او المضلع، لذلك فان كل من مساحات المدرج والمضلع والمنحني هي متساوية. والشكل البياني رقم (9.5) يوضح نموذج لمدرج ومنحني تكراري، علما بان المنحني هو عبارة عن تمهيد Smoothing للمضلع.

شكل بياني رقم (9.5)

يوضح نموذج المدرج والمنحني التكراري



(3) الرسوم والصور البيانية Pictorial presentations

ويستهدف استخدامها الى ايصال المعلومة الى الاشخاص بطريقة مبسطة وكونها اكثر جذبا من ويعتمد شكل الرسوم البيانية على شكل وحدات الظاهرة المعنية كرمز اساس في عرضها لاعطاء صورة تقريبية عن الظاهرة، مع افتراض قيمة محددة للوحدة، فمثلا اذا كنا بصدد عرض تطور وسائط النقل فتكون صورة السيارة كمقياس، واذا اردنا التعبير عن تطور السكان نعتمد صورة تخطيطية للشخص وهكذا .

فالتعبير عن تطور عدد الهواتف الارضية لبلد ما كما هي في سنة 2008 وكان عددها هو 5,455,000 هاتف، وحددنا قيمة 1000000 هاتف لكل صورة فيصبح التعبير عن ذلك كما في الشكل التالي :

☎ = 1000000

☎ = 455000

☎ ☎ ☎ ☎ ☎ ☎

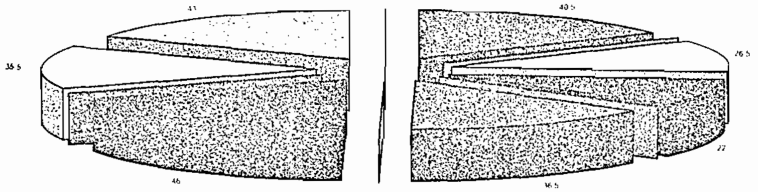
(4) الدائرة البيانية Pie char

واستخدام الدائرة البيانية يستهدف متابعة تطور ظاهرة او متغير معين وابراز الاجزاء التي يتكون منها المتغير كما مبين في الشكل البياني رقم (10.5)، ويتم ذلك من خلال تقسيم مساحة الدائرة الى قطاعات كل منها يمثل جزءا منها. ويتم تحديد كل جزء من خلال ضرب الزاوية المركبة للدائرة والتي مقدارها 360° بحاصل قسمة الجزء المعني على مجموع الاجزاء، اي :

$$\text{مساحة الجزء المعني} = \frac{\text{قيمة مساحة الجزء}}{\text{مجموع قيمة مساحات الاجزاء}} \times 360^\circ$$

شكل بياني رقم (10.5) يمثل نموذج لدائرة بيانية

استخدام برنامج SPSS في المقاييس الوصفية



5-1-6 حالة دراسية رقم $C_{5\text{-}1}$

(1) أستخدام برنامج SPSS في ايجاد المقاييس الوصفية Descriptive Statistics

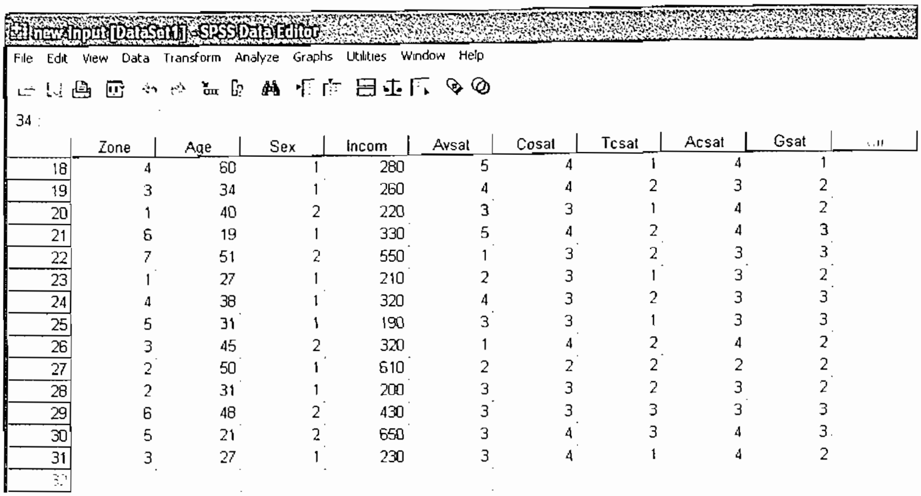
لنفترض كنا بصدد ايجاد المقاييس الوصفية المتعلقة بمؤشرات النزعة المركزية وغير المركزية، ومؤشرات التشتت وخواص الانحراف المعياري، لمعطيات الجدول

رقم (2.1) موضوع الحالة الدراسية C_{1-1} المتعلقة بدراسة استطلاع عينة من مستخدمي النقل العام في عمان للوقوف على مستوى رضاهم عن مستوى خدمات النقل العام. نقوم بمتابعة الخطوات التالية :

- الدخول الى برنامج SPSS، واستدعاء ملف المعطيات الذي تم انشاءه، والمبين في الشكل البياني رقم (11.5)،
- استدعاء القائمة Analysis ومنها الامر الفرعي Descriptive Statistics والكبس على خيار Frequencies، فيظهر لنا مربع الحوار Frequencies المبين في الشكل البياني رقم (12.5) ،
- استخدام السهم الجانبي الموجود على مربع الحوار لنقل المتغيرات الى تحت عنوان Variables، كما مبين على ذات الشكل البياني رقم (12.5) ،
- الكبس على ايقونة Statistics فتظهر لنا لوحة الخيارات : Frequencies Statistics المبينة في الشكل البياني رقم (13.5)، ليتم التاشير على المقاييس الوصفية المطلوبة، ثم الكبس على ايقونة Continue للعودة الى مربع الحوار
- الكبس على ايقونة Charts لتظهر لنا لوحة Frequencies : Charts المبينة في الشكل البياني رقم (14.5) ليتم فيها التاشير على الاشكال البيانية المطلوبة، ومن ثم الكبس على ايقونة Continue للعودة ثانية الى مربع الحوار،
- الكبس على ايقونة Ok فتظهر لنا مخرجات التحليل الاجمالية وكذلك التفصيلية على نطاق كل متغير، وكما مبين نماذج منها في الجداول رقم (1.5) والاشكال البيانية رقم (15.5) .

الشكل البياني رقم (11.5)

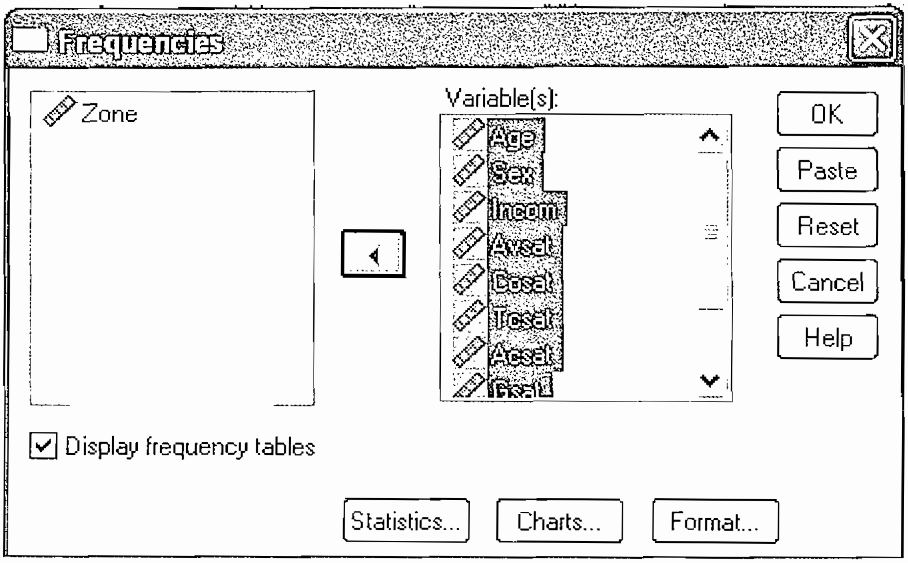
مقطع من ملف الجدول رقم 2.1



	Zone	Age	Sex	Incom	Avsat	Cosat	Tcsat	Acsat	Gsat
18	4	60	1	280	5	4	1	4	1
19	3	34	1	260	4	4	2	3	2
20	1	40	2	220	3	3	1	4	2
21	6	19	1	330	5	4	2	4	3
22	7	51	2	550	1	3	2	3	3
23	1	27	1	210	2	3	1	3	2
24	4	38	1	320	4	3	2	3	3
25	5	31	1	190	3	3	1	3	3
26	3	45	2	320	1	4	2	4	2
27	2	50	1	610	2	2	2	2	2
28	2	31	1	200	3	3	2	3	2
29	6	48	2	430	3	3	3	3	3
30	5	21	2	650	3	4	3	4	3
31	3	27	1	230	3	4	1	4	2

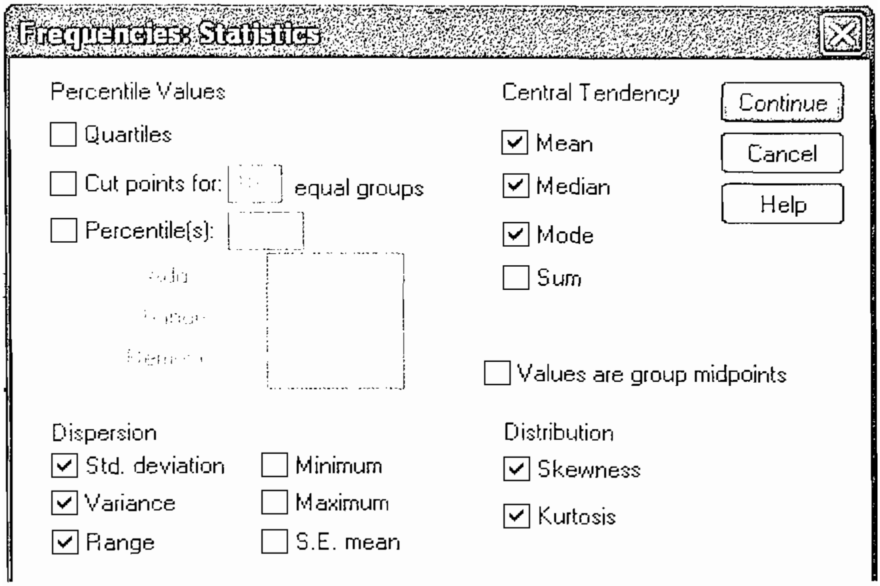
الشكل البياني رقم (12.5)

مربع حوار Frequencies للحصول على مؤشرات وصفية



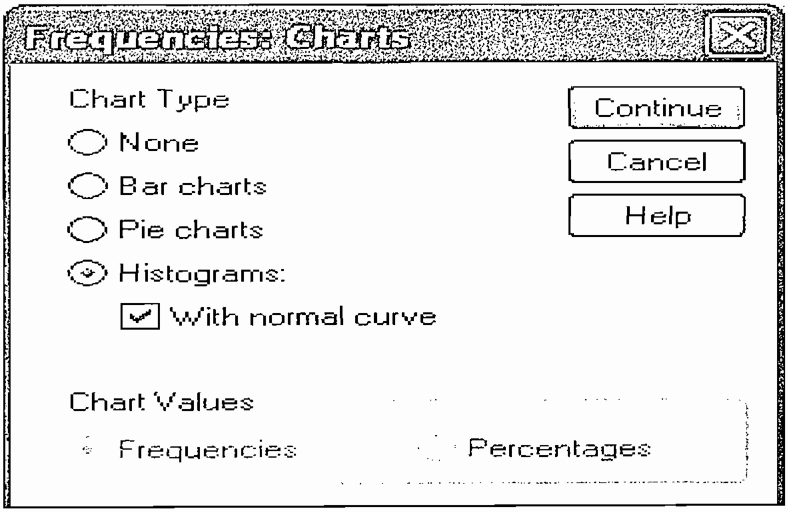
الشكل البياني رقم (13.5)

Frequencies : Statistics لوحة حوار الخيارات



الشكل البياني رقم (14.5)

Frequencies : Charts لوحة حوار



جداول رقم (1.5)

نماذج من مخرجات الخيار Frequencies من المؤشرات الوصفية

		Age	Sex	Incom	Avsat	Cosat	Tcsat	Acsat	Gsat
N	Valid	31	31	31	31	31	31	31	31
	Missing	0	0	0	0	0	0	0	0
	Mean	36.84	1.42	325.48	3.58	3.61	2.06	3.16	2.58
	Median	34.00	1.00	310.00	4.00	4.00	2.00	3.00	3.00
	Mode	34	1	260	3(a)	4	2	3	3
	Std. Deviation	12.833	.502	124.33	1.119	.715	.772	.934	.886
	Variance	164.67	.252	15458.9	1.252	.512	.596	.873	.785
	Skewness	.195	.344	1.143	-.597	.159	.351	-.603	.354
	Std. Error of Skewness	.421	.421	.421	.421	.421	.421	.421	.421
	Kurtosis	-1.174	-2.017	.701	.081	-.213	-.068	.296	.825
	Std. Error of Kurtosis	.821	.821	.821	.821	.821	.821	.821	.821
	Range	41	1	470	4	3	3	4	4
Percentiles	25	25.00	1.00	230.00	3.00	3.00	2.00	3.00	2.00
	50	34.00	1.00	310.00	4.00	4.00	2.00	3.00	3.00
	75	48.00	2.00	400.00	4.00	4.00	3.00	4.00	3.00

a Multiple modes exist. The smallest value is shown

Sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	18	58.1	58.1	58.1
	2	13	41.9	41.9	100.0
	Total	31	100.0	100.0	

Avsat

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	2	6.5	6.5	6.5
	2	2	6.5	6.5	12.9
	3	10	32.3	32.3	45.2
	4	10	32.3	32.3	77.4
	5	7	22.6	22.6	100.0
	Total	31	100.0	100.0	

Tcsat

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	7	22.6	22.6	22.6
	2	16	51.6	51.6	74.2
	3	7	22.6	22.6	96.8
	4	1	3.2	3.2	100.0
	Total	31	100.0	100.0	

Acsat

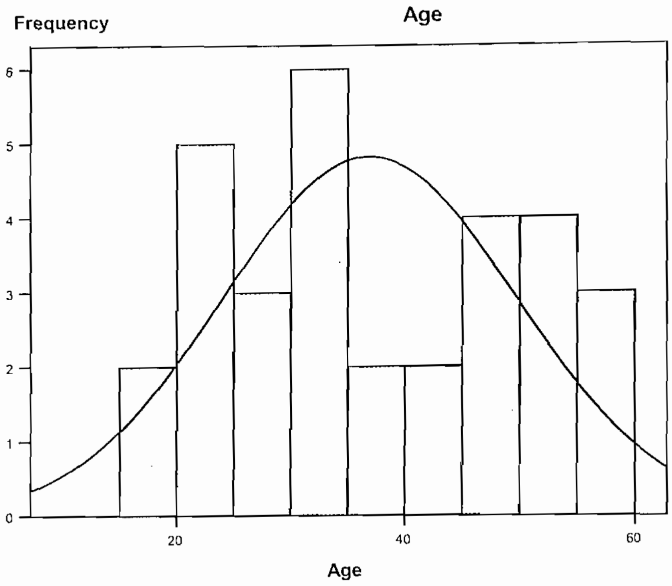
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	2	6.5	6.5	6.5
	2	4	12.9	12.9	19.4
	3	13	41.9	41.9	61.3
	4	11	35.5	35.5	96.8
	5	1	3.2	3.2	100.0
	Total	31	100.0	100.0	

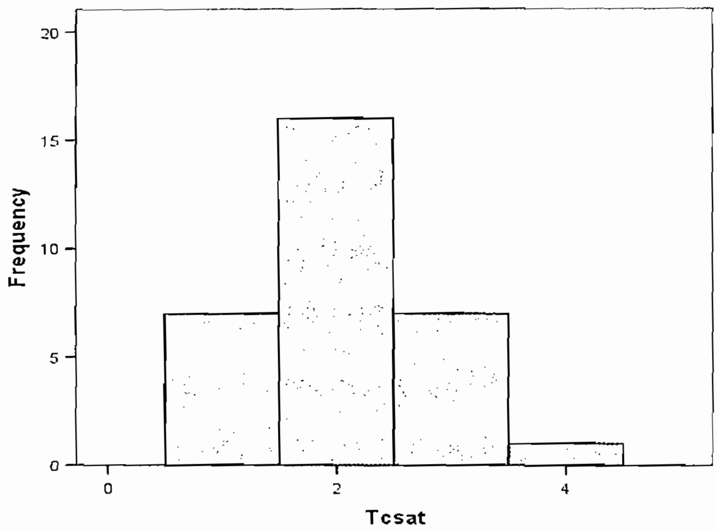
Gsat

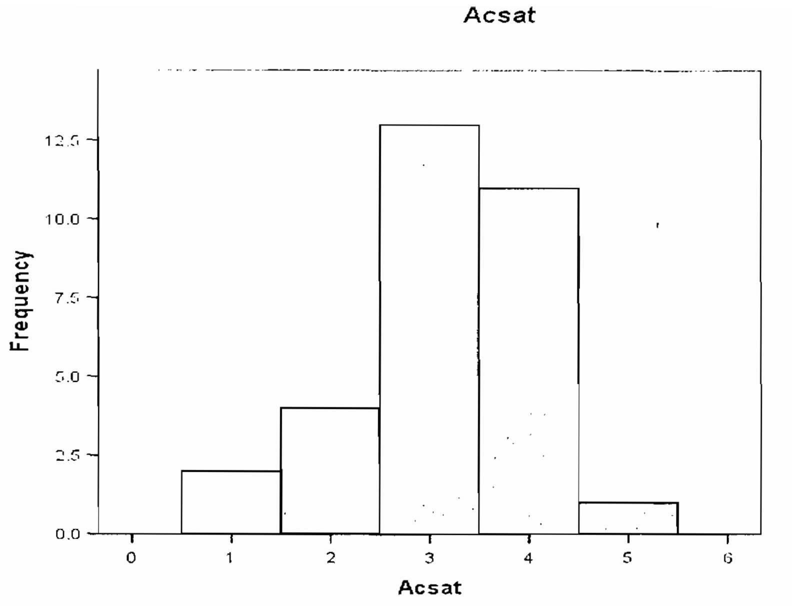
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	3	9.7	9.7	9.7
	2	11	35.5	35.5	45.2
	3	14	45.2	45.2	90.3
	4	2	6.5	6.5	96.8
	5	1	3.2	3.2	100.0
	Total	31	100.0	100.0	

الاشكال البيانية رقم (15.5)

نماذج من مخرجات Charts من الخيار Frequencies







2-5 الارتباط في التحليل الوصفي

Correlation for Descriptive Analysis

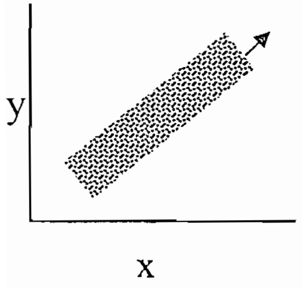
والارتباط ايضا هو من ادوات التحليل الوصفي ويهدف الى معرفة ان كانت هناك علاقة بين متغيرين مستقلين او بين متغير مستقل (x) independent variable ومتغير تابع (y) dependent variable او بين مجموعة متغيرات مستقلة (X_i) ومتغير تابع (y) . بشرط ان يكون كلا المتغيرين عشوائيين وتوزيعهما طبيعيا زوجيا Bivariate normal distribution في حالة زوج من المتغيرات، ويدعى بالتوزيع الطبيعي متعدد المتغيرات Multivariate normal distribution في حالة العلاقة بين مجموعة متغيرات. أما في حالة كان توزيع قيم المتغيرات غير طبيعي فلا يمكن الاستدلال على تقديرمعالم المجتمع من نتائج العينة رغم امكانية احتساب مقياس الارتباط واستخدامه لوصف العلاقة .

ومقياس العلاقة يدعى معامل الارتباط correlation coefficient ويرمز له r في حالة العلاقة بين متغيرين و R في حالة العلاقة بين مجموعة متغيرات مستقلة ومتغير تابع. وتكون قيمة معامل الارتباط 1 عندما تكون العلاقة تامة، وقيمته 0 عندما لاتوجد اية علاقة، وبذلك فان معامل الارتباط يقع بين 0 و 1 أي : ($0 \leq r \leq 1$) .

ويقال ان الارتباط موجبا اذا كانت كل زيادة في المتغير المستقل x تؤدي الى زيادة في المتغير التابع y ويأخذ الاتجاه المبين في الشكل البياني رقم (11.5). ويصبح الارتباط سالبا ويأخذ الاتجاه المبين في الشكل البياني رقم (12.5) اذا كانت الزيادة في قيمة X تؤدي الى نقصان في Y اما في الحالة التي لا تؤدي الزيادة في x الى اي تغير في y فذلك يشير الى عدم وجود اي علاقة بين المتغيرين ويأخذ الشكل البياني رقم (13.5).

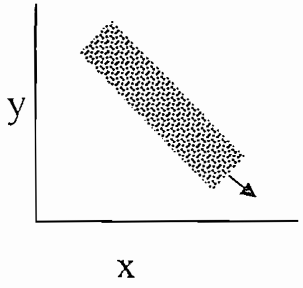
شكل بياني (11.5)

الارتباط موجب



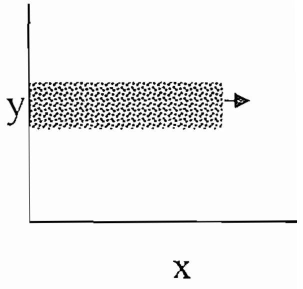
شكل بياني (12.5)

الارتباط سالب



شكل بياني (13.5)

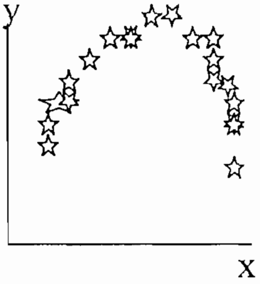
r=0



كما ان شكل الانتشار الذي تؤول اليه العلاقة والمبين نماذج منه في الاشكال البيانية رقم (14.5) و(15.5) و(16.5)، يوضح ان كانت هذه العلاقة هي خطية او غير خطية للاستعانة بها في معرفة الادوات التحليلية المناسب توظيفها في دراسة الظاهرة.

شكل بياني (14.5)

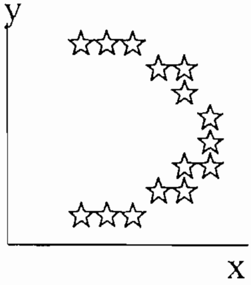
انتشار غير خطي



$E(y) = \alpha + \beta X + CX^2$

شكل بياني (15.5)

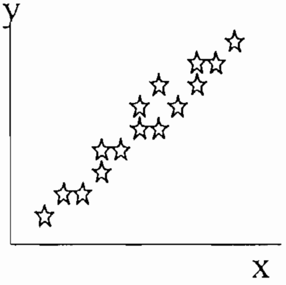
انتشار غير خطي



$E(y) = \alpha \beta^X$

شكل بياني (16.5)

انتشار خطي



$E(y) = \alpha + \beta X$

5- 2- 1 معامل الارتباط البسيط

Simple correlation coeficient

(1) حالة الاستخدام وصيغة حساب معامل الارتباط البسيط

ويستخدم لقياس درجة العلاقة بين متغيرين قيمها كمية quantitative ومن مقاييسه معامل ارتباط بيرسن Pearsons correlation coefficient. وصيغة احتسابه هي :

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\{n\sum x^2 - (\sum x)^2\}\ \{n\sum y^2 - (\sum y)^2\}}}$$

حيث ترمز x و y الى قيم كل من المتغير المستقل والمتغير التابع على التوالي، وتشير n الى حجم العينة .

(2) حالة دراسية $C_{5\text{-}2}$ استخدام برنامج SPSS في تحليل الارتباط البسيط

لنفترض كنا بصدد استخدام معطيات الجدول (2.1)، واستهدفنا معرفة علاقة متغير الرضا Gsat مع كل من متغير: العمرAge، الجنس Sex، الدخل

Incom، ومستوى الرضا عن اجور النقل Tcsat مثلا، فسنحتاج القيام بالخطوات التالية:

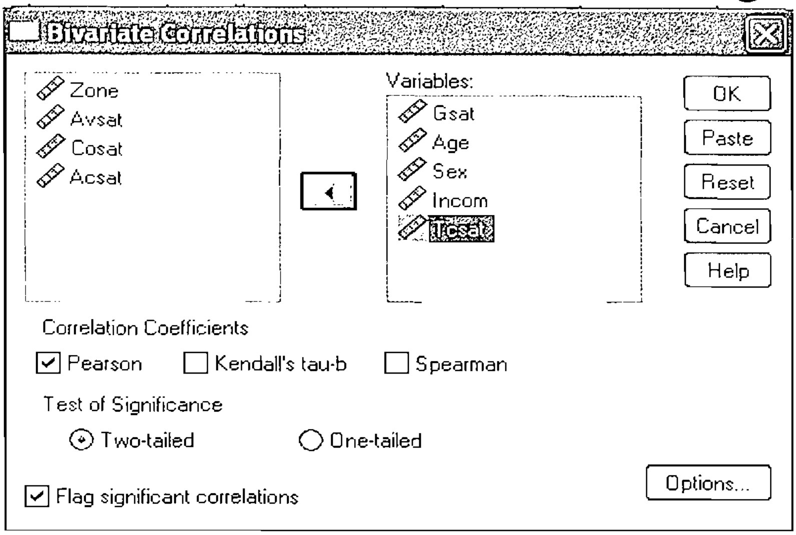
- اخضاع الملف، واستدعاء الامر الفرعي Correltion من قائمة Analysis والكبس على خيارBivariate، فيظهر لنا مربع الحوار المبين في الشكل البياني رقم (17.5)، ليتم استخدام السهم الجانبي لنقل المتغيرات المستهدف ايجاد الارتباط البسيط بين كل منها الى تحت عنوان Variables ،
- الكبس على Ok فنحصل على جدول المخرجات رقم (2.5) .

ومن جدول مخرجات التحليل رقم (2.5) نستدل بان :

- هناك علاقة عالية المعنوية بين المتغير التابع Gsat ومستوى الرضا عن اجور النقل Tcsat، اي عند مستوى معنوية مقدارها $\alpha = 0.01$، كما ونستدل على ان اتجاه العلاقة موجبة، وهذا يعني انه كلما ازداد الرضا عن مستوى الاجور Tcsa من خلال انخفاضها، كلما ازداد رضا مستخدمي النقل العام Gsat مما يدلل على صحة الاشارة لمنطقية الحالة.
- ان هناك علاقة معنوية بين المتغير التابع Gsat ومتغير العمرAge، اي عند مستوى معنوية مقدارها $\alpha = 0.05$، وان اشارة معامل الارتباط جاءت باشارة سالبة، بمعنى كلما كان عمرAge مستخدم النقل العام اصغر، كلما ارتفع مستوى رضاه عن مستوى الخدمات، وهو امر منطقي لان الاصغر عمرا اقل اكتراثا بالاجور الذي يمثل الاعلى علاقة بمستوى الرضا العام عن مستوى خدمات النقل العام .
- اما باقي المتغيرات التي ضمها الجدول فهي لاتدل على علاقة معنوية مع المتغير التابع Gsat .

الشكل البياني رقم (17.5)

مربع حوار الارتباط البسيط Pearsonain Correlation



جداول رقم (2.5)

مخرجات تحليل الارتباط البسيط Bivariate Correlation

		Gsat	Age	Sex	Incom	Tcsat
Gsat	Pearson Correlation	1	-.405(*)	.109	.064	.626(**)
	Sig. (2-tailed)		.024	.560	.733	.000
	N	31	31	31	31	31
Age	Pearson Correlation	-.405(*)	1	.032	.115	-.285
	Sig. (2-tailed)	.024		.866	.539	.120
	N	31	31	31	31	31
Sex	Pearson Correlation	.109	.032	1	.213	.100
	Sig. (2-tailed)	.560	.866		.250	.593
	N	31	31	31	31	31
Incom	Pearson Correlation	.064	.115	.213	1	.271
	Sig. (2-tailed)	.733	.539	.250		.141
	N	31	31	31	31	31
Tcsat	Pearson Correlation	.626(**)	-.285	.100	.271	1
	Sig. (2-tailed)	.000	.120	.593	.141	
	N	31	31	31	31	31

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

5- 2- 2 معامل الارتباط المتعدد
Multiple correlation coefficient

(1) حالة الاستخدام وصيغة حساب معامل الارتباط المتعد

ويستخدم لقياس العلاقة بين اكثر من متغيرين، الا ان اشارة معامل الارتباط لا تدل على اتجاه العلاقة هنا لان هذا الاتجاه لا يكون موحدا لجميع المتغيرات. وصيغة حسابه في حالة 3 متغيرات لايجاد العلاقة بين x_1 و كل من x_2 x_3 هي:

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \quad r_{13} \quad r_{23}}{1 - r_{23}^2}}$$

حيث ان: r_{23} , r_{13} , r_{12} هي معاملات ارتباط يتم ايجادها بموحب صيغة الارتباط البسيط المينة في اعلاه .

(2) حالة دراسية رقم $C_{5\text{-}3}$ استخدام برنامج SPSS

ان موضوع الارتباط المتعدد يرتبط بموضوع الانحدار لانه يبحث في علاقة وتاثير المتغيرات المستقلة x_i على المتغير التابع y، وان هذه العلاقة تقوم على اساس انها خطية. لذا فان قيم كل من R و R^2 ونتائج اختبار معنويتها باستخدام *f* هي من ضمن ما تشتمل عليه مخرجات تحليل الانحدار Regression Analysis.

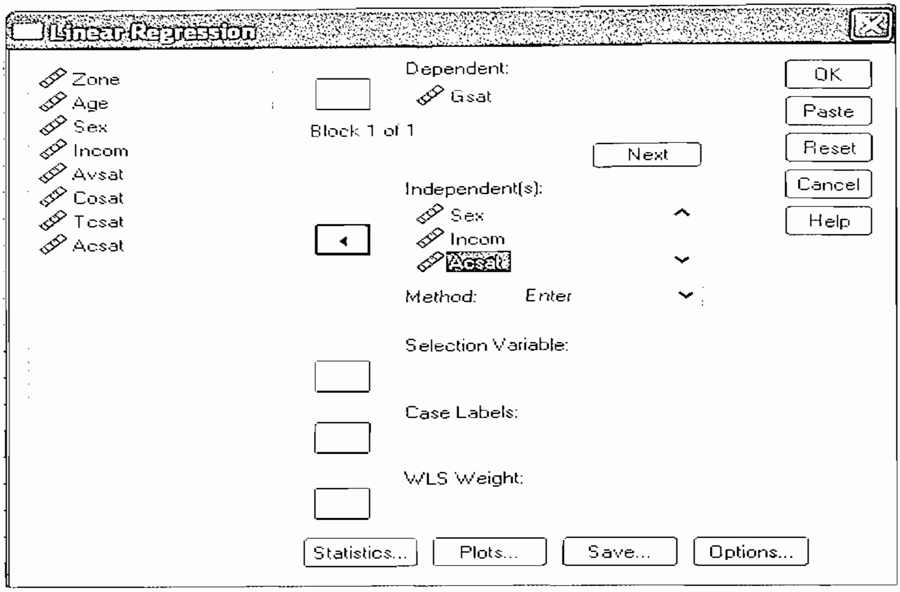
وحيث سيتم لاحقا تناول موضوع الانحدار بصورة مفصلة لاهميته، لذا سيتم التطرق هنا بقدر ما يتعلق الامر بايجاد الارتباط المتعدد لاكثر من متغيرين باستخدام برنامج SPSS، من خلال توظيف معطيات الجدول رقم (2.1) موضوع المثال (1.1)، والتي يمكن تلحيص اجراءات الحصول عليه بما يلي :

- استدعاء القائمة Analysis ومنها الامر الفرعي Regression، ثم الخيارLincar ، فيظهر لنا مربع الحوار Linear Regression المبين في الشكل البياني رقم (18.5)، وفيه يتم استخدام السهم الجانبي لنقل المتغيراتTcsat, Avsat, Acsat, Sex, Age, Cosat, Incom الى تحت Independents و المتغير التابع Gsat الى تحت Dependent،
- الكبس على ايقونة Statistics ، فتظهر لنا لوحة Linear Regression Statistics : المبينة في الشكل البياني رقم (19.5) ليتم التاشير فيها على R، ثم الكبس على ايقونة Continue للعودة الى مربع الحوار Linear Rcgression مرة اخرى ،
- الكبس ايضا على ايقونة Options لنحصل عل لوحة Options : Linear Regression المبينة في الشكل البياني رقم (20.5) ليتم التاشير فيها على اختبار f والعودة من جديد الى مربع الحوار،
- الكبس على ايقونة Ok لنحصل على المخرجات المينة في الجداول رقم (3.5).

ومن جدول المخرجات (3.5) نستدل على ان معامل الارتباط المتعدة R مقداره 0.678، وان قيمة f جاءت معنوية عند $\alpha = 0.04$ وبدرجات حرية عددها 2 و 3، عليه نرفض فرضية H_0 مما يدل على ان معامل ارتباط المجتمع المتعدد لايساوي صفر.

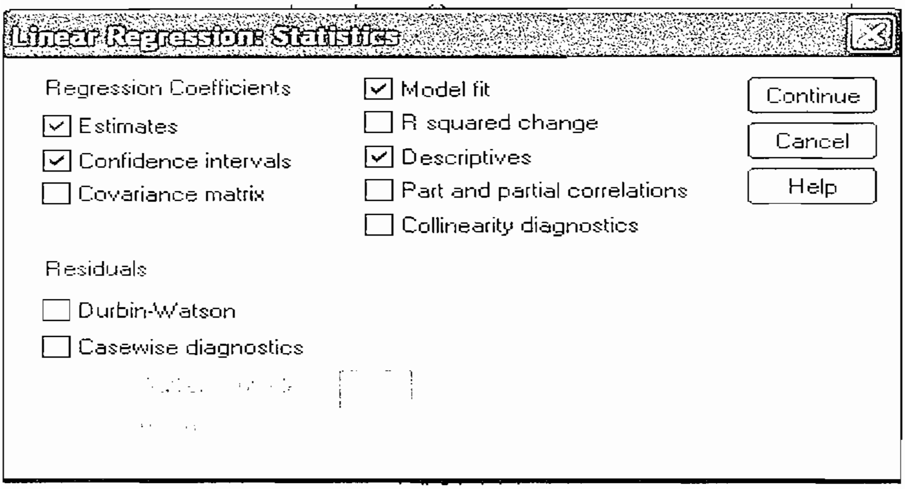
الشكل البياني رقم (18.5)

مربع الحوار Linear Regression لايجاد معامل الارتباط المتعدد



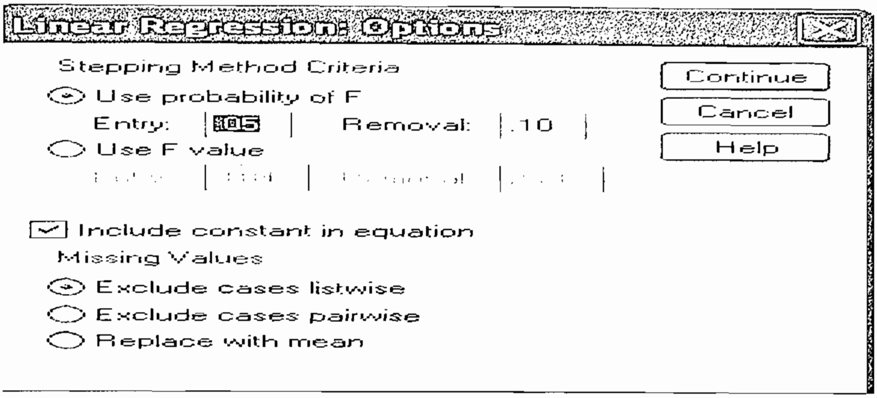
الشكل البياني رقم (19.5)

لوحة Linear Regression : Statistics لمعامل الارتباط المتعدد



الشكل البياني رقم (20.5)

لوحة Linear Regression : Options لمعامل الارتباط المتعدد



جداول مخرجات رقم (3.5)

مخرجات الانحدار الخطي لايجاد الارتباط المتعدد

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	Tcsat, Avsat, Acsat, Sex, Age, Cosat, Incom(a)	.	Enter

a All requested variables entered.

b Dependent Variable: Gsat

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.678(a)	.460	.295	.744

a Predictors: (Constant), Tcsat, Avsat, Acsat, Sex, Age, Cosat, Incom

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10.823	7	1.546	2.795	.029(a)
	Residual	12.725	23	.553		
	Total	23.548	30			

a Predictors: (Constant), Tcsat, Avsat, Acsat, Sex, Age, Cosat, Incom

b Dependent Variable: Gsat

5- 2- 3 معامل الاتباط الجزئي artial correlation coefficient

(1) حالة الاستخدام وصيغة حساب معامل الارتباط الجزئي

ويستخدم لقياس العلاقة بين زوج من المتغيرات عندما باقي المتغيرات تكون ثابتة. وبذلك فان الفرق بين الارتباط البسيط والارتباط الجزئي هو ان الاول يقيس العلاقة بين متغيرين ضمن تاثيرالمتغيرات الاخرى، في حين يقيس الثاني العلاقة بين متغيرين مع استبعاد تاثير المتغيرات الاخرى. وصيغة حساب معامل الارتباط الجزئي بين y و x_2 مع ثبات x_1 مثلا هي :

$$r_{y2.1} = \frac{r_{y2} - (r_{y1})\ (r_{12})}{\sqrt{(1 - r_{y1}^2)\ (1 - r_{12}^2)}}$$

حيث ان r_{y2} و r_{y1} و r_{12} هي معاملات يتم ايجادها بموحب صيغة الارتباط بسيط.

(2) حالة دراسية رقم $C_{5\text{-}4}$ استخدام برنامج SPSS في تحليل الارتباط المتعدد

لنفترض كنا بصدد ايجاد العلاقة بين المتغيرات Age, Tcsat, Gsat مع ثبات تاثير المنغير Age، موضوع الجدول رقم (2.1)، فان الاجراءات المطلوبة للحصول على مخرجات التحليل تتمثل بالخطوات التالية :

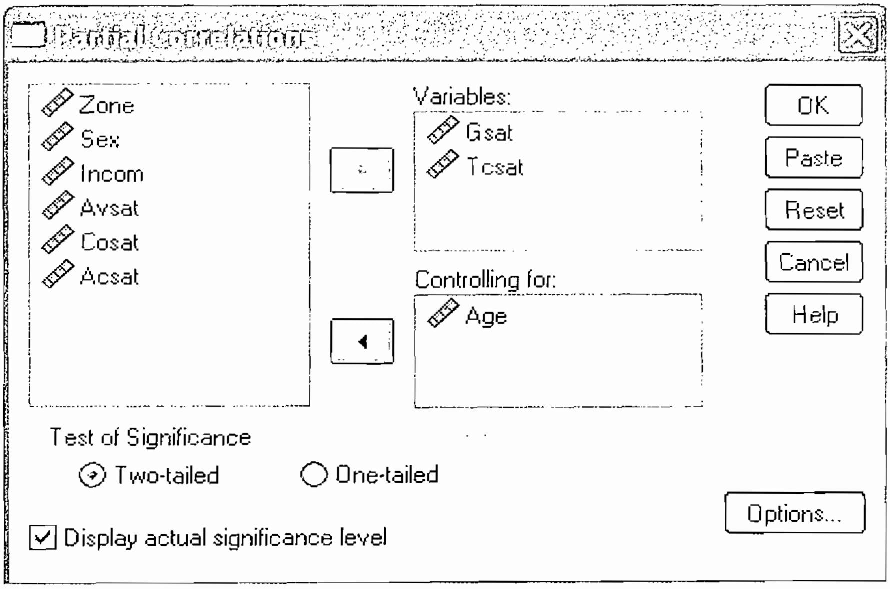
- استدعاء القائمة Analysis ومنها الامر الفرعي Correlate ثم الخيار Partial فنحصل على مربع الحوار Partial Correlation المبين في الشكل البياني رقم (21.5). وفيه يتم استخدام السهم الجانبي لنقل المتغيرين Tcsat, Gsat الى تحت Variables و Age تحت Controlling for والتاشير عند مستوى المعنوية المطلوبة تحت Significance Test of،

- ثم الكبس على ايقونة Options لتظهر لنا لوحة Partial Correlation : Options المبينة في الشكل البياني رقم (22.5) ليتم عليها التاشير على means and standard deviations تحت Statistics، وبعدها الكبس على ايقونة Continue للعودة مربع حوار Partial Correlations ،
- الكبس على ايقونة Ok لنحصل على المخرجات المبينة في الجداول رقم (4.5).

ومن المخرجات نستدل على ان معامل الارتباط الجزئي هو 0.582 وباشارة موجبة ايضا ، الا ان فرضية H_0 مرفوضة، اي ان قيمة معامل الارتباط لاتساوي صفر، وان مستوى المعنوية عالية حيث ان $\alpha = 0.001$.

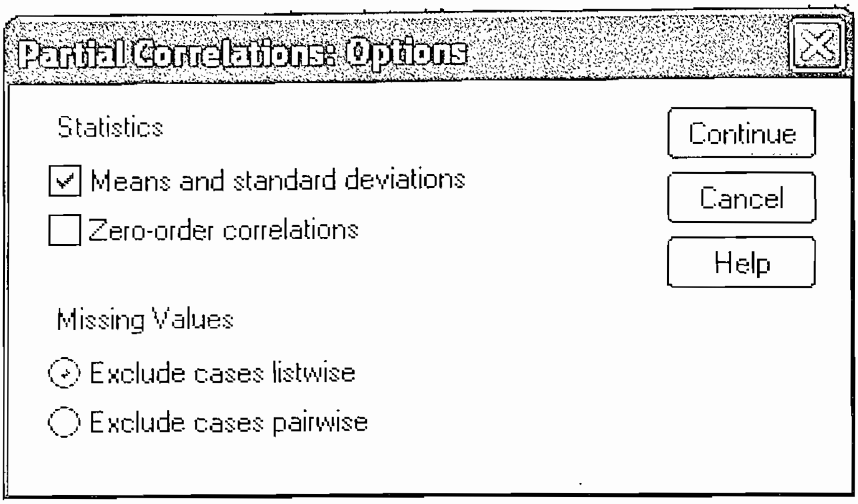
الشكل البياني رقم (21.5)

مربع حوار Partial Correlation



الشكل البياني رقم (22.5)

لوحة **Partial Correlation : Options**



جداول رقم (4.5)

مخرجات معامل الارتباط الجزئي

Descriptive Statistics

	Mean	Std. Deviation	N
Gsat	2.58	.886	31
Tcsat	2.06	.772	31
Age	36.84	12.833	31

Correlations

Control Variables			Gsat	Tcsat
Age	Gsat	Correlation	1.000	.582
		Significance (2-tailed)	.	.001
		df	0	28
	Tcsat	Correlation	.582	1.000
		Significance (2-tailed)	.001	.
		df	28	0

4-2-5 معامل ارتباط الرتب Rank correlation coefficient

(1) حالة الاستخدام وصيغة حساب معامل ارتباط الرتب

ومن مقاييسه معامل ارتباط سبيرمان (1)Spearman rank correlation coefficient ويمكن استخدامه مع كل من المعطيات الكمية quantitative والمعطيات النوعية qualtitative ويعود الى فصيلة التوزيعات الحرة (غير المعلمية) اي التي لايشترط فيها الاستيفاء بشرط التوزيع الطبيعي لقيم متغيراتها، ويفضل استخدامه مع المعطيات النوعية فقط لان استخدامه مع المعطيات الكمية يكون اقل دقة من معامل ارتباط بيرسون البسيط، وصيغته هي :

$$r_s = \frac{6\sum d_i^2}{n(n^2-1)}$$

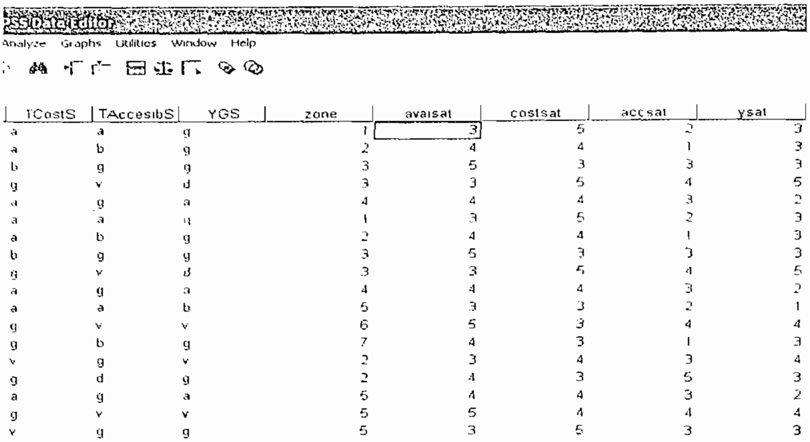
حيث ان d هو الفرق بين رتبة او تسلسل مشاهده ما حسب المتغير الاول X_1 ورتبتها حسب المتغير الثاني X_2. وعندما يكون هناك عدة مشاهدات بنفس المستوى يعتبر الوسط الحسابي هو رتبة كل واحدة من تلك المشاهدات عند رتبتها تصاعديا. وان n هي عدد المشاهدات .

(2) حالة دراسية رقم استخدام برنامج SPSS في تحليل ارتباط الرتب

نتابع في الاتي ايجاد العلاقة بين متغيري مستوى الرضا العام YGS ومستوى الرضا عن اجور النقل TCostS موضوع الجدول (1.1) اللذان كلاهما قيم نوعية وتم تحويلهما الى قيم كمية باستخدام الخيار Recode من القائمة Transform واللذان اعيد تسميتهما بـ ysat و costsat على التوالي والمبين مقطع من الملف في الشكل البياني رقم (23.5) ادناه :

الشكل البياني رقم (23.5)

مقطع من ملف المعطيات قبل وبعد تحويل القيم من نوعية الى كمية



TCostS	TAccesibS	YGS	zone	avaisat	costsat	accsat	ysat
a	a	g	1	3	5	2	3
a	b	g	2	4	4	1	3
b	g	g	3	5	3	3	3
g	v	d	3	3	5	4	5
a	g	a	4	4	4	3	2
a	a	g	1	3	5	2	3
a	b	g	2	4	4	1	3
b	g	g	3	5	3	3	3
g	v	d	3	3	5	4	5
a	g	a	4	4	4	3	2
a	a	b	5	3	3	2	1
g	v	v	6	5	3	4	4
g	b	g	7	4	3	1	3
v	g	v	2	3	4	3	4
g	d	g	2	4	3	5	3
a	g	a	5	4	4	3	2
g	v	v	5	5	4	4	4
v	g	g	5	3	5	3	3

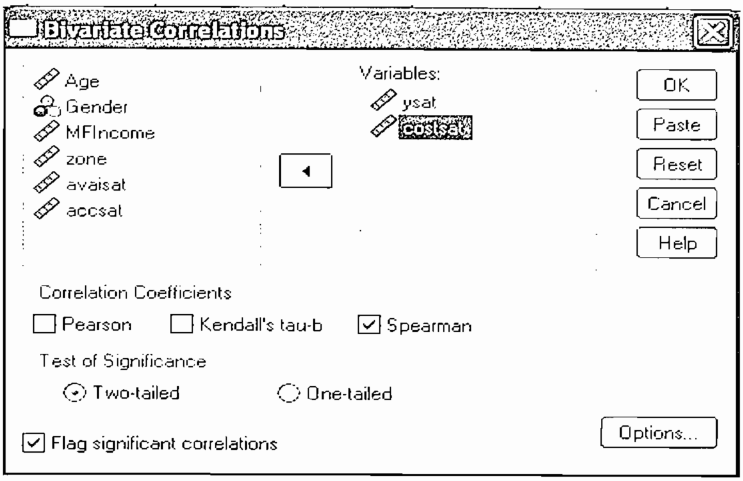
اما الخطوات الاخرى المطلوبة للحصول على معامل ارتباط الرتب فهي لاتختلف عن تلك المتعلقة بمعامل الارتباط التنائي البسيط او الجزئي، باستثناء التاشير على Spearman تحت Correlation Coefficient بدلا من Pearson او Partial اي :

- استدعاء القائمة Analysis ومنها الامر الفرعي Correlate ومن ثم الخيار Bivariate ليظهر مربع الحوارBivariate Correlations المبين في الشكل البياني رقم (24.5)، لينم التاشيرعلى Spearman تحت Correlation Coefficient، واستخدام السهم الجانبي لنقل المتغيرين الى تحت Variables.
- وبعد التعامل مع ايقونة Options والعودة الى مربع الحوار، يتم الكبس على ايقونة Ok لنحصل على مخرجات التحليل المبينة في الجداول رقم (5.5) .

ومن المخرجات نستدل على ضعف معامل الارتباط . مما يدل على عدم وجود علاقة معنوية بين المتغيرين وكما يعزز ذلك قيمة مستوى المعنوية Sig. at 0.317.

الشكل البياني رقم (24.5)

يوضح التاشيرعند Spearman للحصول على معامل ارتباط الرتب



جداول رقم (5.5)

مخرجات استخدام برنامج SPSS في الحصول

على معامل ارتباط الرتب بين متغيري ysat و costsat

			ysat	costsat
Spearman's rho	ysat	Correlation Coefficient	1.000	.186
		Sig. (2-tailed)	.	.317
		N	31	31
	costsat	Correlation Coefficient	.186	1.000
		Sig. (2-tailed)	.317	.
		N	31	31

5-2-5 معامل ارتباط الاقتران

Association correlation coefficient

ويستخدم في الحالات التي يكون فيها معطيات كلا المتغيرين او احدهما غير قابلة للترتيب التصاعدي او التنازلي، وان كل من المتغيرين يتكون من مستويين (حالتين). وصيغة حسابه وفقا لجدول المعطيات ادناه هي:

variable	1	2
A	n_{a1}	n_{a2}
b	n_{b1}	n_{b2}

$$r_A = \frac{n_{a1}n_{b2} - n_{a2}n_{b2}}{n_{a1}n_{b2} + n_{a2}n_{b1}}$$

6-2-5 معامل ارتباط التوافق

Contingency correlation coefficient

(1) حالة الاستخدام وصيغة حساب معامل ارتباط التوافق

ويستهدف قياس العلاقة بين متغيرين يكون احدهما او كلاهما ينقسم الى اكثر من حالتين. وان الشكل العام لجدول التوافق في عرض المتغيرين هو كما مبين فالجدول رقم (1.5) التالي :

جدول رقم (5.5)

يبين الشكل العام لصيغة عرض المتغيرين في حالة ارتباط التوافق

المتغير الثاني	المتغير الاول X_i				
y_i	X_1	X_2	....	X_c	المجموع
y_1	$n_{y1\,x1}$	$n_{y1\,x2}$	....	$n_{y1\,xc}$	n_{y1}
y_2	$n_{y2\,x1}$	$n_{y2\,x2}$	....	$n_{y2\,xc}$	n_{y2}
. . .	. . .	. . .		. . .	. . .
y_r	$n_{yr\,x1}$	$n_{yr\,x2}$	....	$n_{yr\,xc}$	n_{yr}
المجموع	n_{x1}	n_{x2}	....	n_{xc}	n

وصيغة حسابه هي :

$$r_C = \sqrt{\frac{\chi_c^2}{\chi_c^2 + 2}}$$

حيث ان :

$$\chi_c^2 = n\left[\frac{n_{y1x1}^2}{n_{y1}n_{x1}} + \frac{n_{y1x2}^2}{n_{y1}n_{x2}} + \ldots + \frac{n_{y1XC}^2}{n_{yr}n_{XC}}\right] - n$$

(1) حالة دراسية رقم C5-6

استخدام برنامج SPSS في اختبار التجانس، χ^2 لانجاز تحليل ارتباط التوافق .

ويمكن انجاز تحليل ارتباط التوافق على مرحلتين، يتم في المرحلة الاولى ايجاد قيمة χ^2 باستخدام الخيار Chi-square من الامر الفرعي Non-parametric test في قائمة Analysis، ويجري في المرحلة الثانية حساب معامل ارتباط التوافق من خلال الصيغة :

$$r_C = \sqrt{\frac{\chi_c^2}{\chi_c^2 + 2}}$$

في دراسة قامت بها قناة تلفزيونية لمعرفة كان برنامجها الترفيهي له نفس الاهتمام بين كافة الفئات العمرية، فاختارت عينة من المشاهدين حجمها n = 74 وحصلت على النتائج المبينة في الجدول رقم (6.5)، والمطلوب استخدام برنامج SPSS لاختبار ان كان هناك فروق في رغبة مشاهدة البرنامج بين الفئات العمرية عند مستوى معنوية $\alpha = 0.05$.

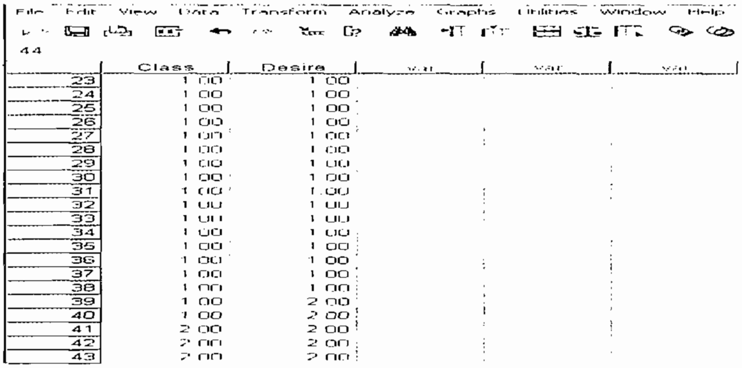
جدول رقم (6.5) : عينة من مشاهدي قناة تلفزيونية حسب الفئة العمرية والرغبة في مشاهدة البرنامج الترفيهي

الفئة العمرية	مستوى الرغبة			
	لايرغب	يرغب	يرغب جدا	المجموع
أقل من 18	20	16	4	40
18- 50	10	8	3	21
50 فاكثر	6	5	2	13
المجموع	36	29	9	74

- من المفيد الاشارة اولا الى ان تكوين الملف لاستخدام برنامج SPSS في اختبار التجانس يتطلب اعطاء المتغير الاول وهي الفئات العمرية القيم 1 للفئة الاولى والقيم 2 للثانية وتاخذ الفئة الثالثة القيم 3 , وعلى نفس الغرار بالنسبة للمتغير الثاني وهو متغيرالرغبة، تعطى القيم 1 لحالة عدم الرغبة والقيم 2 للرغبة والقيم 3 لحالة راغب جدا، وهذا طبعا لكل قيمة من قيم المعطيات اي لغلية 40 قيمة للفئة الاولى ولغاية 36 قيمة لحالة عدم الرغبة من المتغير الثاني وهكذا، وكما مبين في الشكل البياني رقم (25.5) .

الشكل البياني رقم (25.5) :

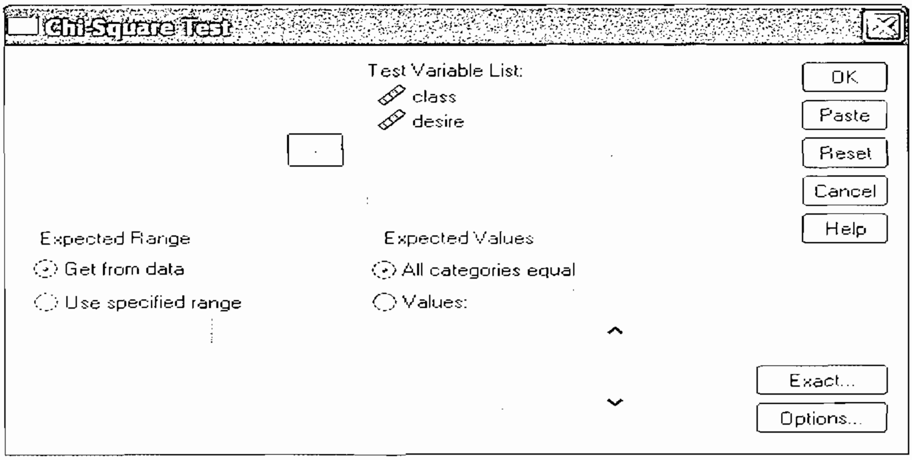
اسلوب ادخال المعطيات لتكوين ملف لاختبار χ^2



يتم اخضاع الملف للامر Analysis ومنه الامر الفرعي Non-parametric test ثم الكبس على الخيار Chi-square ليظهر مربع الحوار المبين في الشكل البياني رقم (26.5) ،

الشكل البياني رقم (26.5)

مربع حوار اختبار التجانس باستخدام مربعات كاي χ^2



- يتم نقل المتغيرين تحت عنوان Test Variable List باستخدام السهم الموجود بجنب مربع الحوار، ومن ثم التاشير عند All Categories Equal،
- الكبس على ايقونة Ok فنحصل على المخرجات المبينة في الجداول رقم (7.5) بضمنها جدول الاحصاء الوصفي الذي يشير الى تشابه قيمتي متوسطي المتغيرين والى تجانس الاراء ضمن الفئات العمرية كما يتضح من قيم الانحراف المعياري لكلا المتغيرين، كما و يستدل على معنوية النتائج عند درجة ثقة 95 ٪ التي جاءت عند درجة معنوية 0.000، حيث ان معنوية asymptotic significance التي تعتمد على توزيع asymp. Distribution تعتبر مقبولة عند اقل من 5 ٪. اي قبول H_0 القائلة بتجانس معايير التصنيف لكلا المتغيرين .

جداول رقم (7.5)

مخرجات برنامج SPSS لاستخدام اختبار Chi-Square Test

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
age groups	74	1.6351	.7688	1.00	3.00
level of wish	74	1.6351	.6939	1.00	3.00

age groups

	Observed N	Expected N	Residual
1.00	40	24.7	15.3
2.00	21	24.7	-3.7
3.00	13	24.7	-11.7
Total	74		

level of wish

	Observed N	Expected N	Residual
1.00	36	24.7	11.3
2.00	29	24.7	4.3
3.00	9	24.7	-15.7
Total	74		

Test Statistics

	age groups	level of wish
Chi-Square[a]	15.595	15.919
df	2	2
Asymp. Sig.	.000	.000

من الجدول اعلاه لدينا :

$$\chi_c^2 = n(\chi^2) - n = 74(15.919) - 74 = 1050.13$$

وبتطبيق صيغة حساب معامل ارتباط التوافق نحصل على :

$$r_C = \sqrt{\frac{\chi_c^2}{\chi_c^2 + 2}} = \sqrt{\frac{1050.13}{1050.13 + 74}} = \sqrt{0.934} = 0.966$$

5- 3 استخدام الانحدار في التحليل الوصفي

Regeression for Descriptive Analysis

5-3-1 خصائص استخدام الانحدار في التحليل الوصفي

تستهدف عملية توظيف الانحدار في التحليل الوصفي هو لاعطاء صورة عن ظاهرة ما والعوامل المحيطة بها اكثر منه الحصول على نتائج معنوية خاضعة لمعايير احصائية وفرضيات علمية، حيث المطلوب هو تضمين كافة المتغيرات في المعادلة من دون الحاجة الى التحقق من كون بعض هذه المتغيرات او اغلبها غير مستوفية للمعايير والفرضيات او من دون ان يكون لها دور فعلي في التاثير على الظاهرة التي تكون موضوع االدراسة، هذا بالاضافة الى ارتفاع كلفتها نتيجة الحاجة الى قياس عدد كبيرمن المتغيرات. وبهذا يكون من الطبيعي الحصول على معامل ارتباط متعدد R عالي المعنوية ونسبة تفسير للتباين ممثلا بمعامل التحديد R^2 عالية ايضا واحيانا شبه تامتين وذلك بسبب الترابط شبه التام بين بعض المتغيرات التي يتضمنها النموذج، ولذلك فهي نماذج غير صالحة لاستخدامها في بناء توقعات او تقديرات او لاغراض السيطرة والتحكم لانها تؤدي الى نتائج غير موثوقه .

5- 3- 2 حالة دراسية رقم $C_{5\text{-}7}$

استخدام برنامج SPSS بتوطيف الانحدار في التحليل الوصفي

في دراسة تستهدف وصف ظاهرة العمل البحثى في الجامعات من خلال عوامل يعتقد بان لها تاثيرعلى الظاهرة وفقا لاراء شريحة من التدريسيين العاملين في جامعات عربية، والوقوف على مدى رضاهم عن ظروف النشاط البحثي وتطويره. شملت عينة عددها n= 74 تدريسي من جامعات

عراقية واردنية واماراتية ويمنية. وشملت المعطيات ايضا معلومات اكاديمية واقتصادية وشخصية عن المبحوثين. وتم تقسيم مستوى الرضا الى 5 درجات، اعطيت بموجبها القيمة 5 لمستوى الرضا التام نزولا عند القيمة 1 لمستوى الرضا الضعيف، والمتغيرات التي تم اخضاعها لتحليل الانحدار تشمل :

1) متغير الجنس Sex (ذكر = 1، انثى = 2)
2) متغير العمر Age (24-35 = 1، 36-45 = 2، 46 -55 = 3، 56 سنة فاكثر = 4)
3) متغير التحصيل العلمي للباحثين الاكاديميين Degree (دكتوراه = 1، ماجستير = 2)
4) متغير بلد التخرج لاخر شهادة Cou(بريطانيا =1، امريكا=2، اخرى=3)
5) متغير اللقب الاكاديمي للباحث Tit (استاذ= 1، استاذ مشارك= 2، استاذ مساعد= 3، اخرى= 4)
6) متغير عدد سنين الخدمة الاكاديمية Ays
7) متغير اجمالي عدد سنين الخدمة الاكاديمية والعملية Tys (1-5 =1، 6-11 = 2، 12 -17 = 3، 18 سنة فاكثر = 4)
8) متغير البلد الذي تقع فيه الجامعة التي يعمل بها الباحث Uni (العراق =1، الاردن = 2، اليمن = 3، دولة الامارات = 4)
9) متغير تخصصات التدريسيين Spe (علوم=1، انسانية وادبية=2، ادارة ومحاسبة=3، علوم طبية=4، اختصاصات هندسية=5، حاسبات وتكنولوجيا=6)
10) متغير مستوى الرضا عن الراتب الشهري X01

11)متغير مستوى الرضا عن ظروف العمل المكانية X_{02}

12)متغير مستوى الرضا عن توفر الاجهزة والمراجع والدوريات X_{03}

13)متغير مستوى الرضا عن العلاقات العامة في العمل X_{04}

14)متغير مستوى الرضا عن حصول الباحث على حقوق الترقية X_{05}

15)متغير مستوى الرضا عن فرص الايفاد وحضور المؤتمرات X_{06}

16)متغير مستوى الرضا عن الموازنة بين ساعات العمل البحثي و المحاضرات X_{07}

17)متغير مستوى الرضا عن الاهتمام بالكفاءات العلمية X_{08}

18)متغير مستوى الرضا عن الاهتمام بالمبادرات التطويرية التي يطرحها الباحثون X_{09}

19)متغيرمستوى الرضا عن المحفزات المادية والمعنوية للباحثين X_{10}

20)متغيرمستوى الرضا عن جدية جهة العمل بالبحوث والدراسات X_{11}

21)متغيرمستوى الرضا عن قوانين وتعليمات البحث العلمي X_{12}

22)متغيرمستوى الرضا عن تعاون المؤسسات الاخرى في توفير المعطيات X_{13}

23)متغيرمستوى الرضا عن جدية تطبيق نتائج البحوث والدراسات X_{14}

24)متغيرمستوى الرضا عن كفاية الدوريات التي تتولى نشر البحوث X_{15}

25)المتغير التابع (عدد الكتب والبحوث المنشورة) y

ان توظيف الانحدار في حالة التحليل الوصفي يعني تضمين كافة المتغيرات اعلاه في معادلة الانحدارسواء أكانت معنوية او غير معنوية، وباستخدام برنامج SPSS في عملية التحليل وبقدر تعلق الامر بالوصف والتفسير نحتاج الى متابعة الخطوات التالية :

- انشاء ملف بالمعطيات اعلاه، بتسمية المتغيرات في صفحة Variable View ونقل معطيات المتغيرين او تدوينهما على صفحة Variable Data وكما مبين في الشكل البياني رقم (27.5) .

- استدعاء قائمة Analysis ومنها الامر الفرعي Regression ومن ثم التاشير على خيار Linear

- يظهر لنا مربع الحوار Linear Regression المبين في الشكل البياني رقم (28.5)، وفيه يتم استحدام السهم الجانبي لنقل المتغير التابع y الى تحت Dependent والمتغيرات المستقلة الى تحت Independent،

- الكبس على ايقونة Statistics لتظهر لنا لوحة Linear Regression: Statistics المبينة في الشكل البياني رقم (29.5) ليتم التاشير ازاء المعايير الوصفية المتعلقة بقياس معنوية النموذج ومعاملات النموذج وكما هو موضح على الشكل البياني. بعد الانتهاء مع لوحة Linear Regression: Statistics والكبس على ايقونة Continue للعودة الى مربع الحوار من جديد

- الكبس على ايقونة Options للحصول على لوحة Linear Regression: Options للتاشير عندها في حالة الرغبة في تغييرما هو مثبت من معايير ادخال المتغير للتحليل او حذفه. والعودة مرة اخرى الى مربع الحوار، وكذا التكرار مع ايقونة Plots في حالة الرغبة او الحاجة في الحصول على اشكال بيانية ،

- الكبس على ايقونة Ok لنحصل على المخرجات المبينة في الجداول رقم (8.5).

ومن المخرجات نستدل على ان جميع المعايير الاحصائية وهي، F, R , R^2 عالية المعنوية عند $\alpha = 0.000$، ولكن هذه المعايير العالية المعنوية غير كافية لوحدها لصلاحية النموذج لاستخدامة في بناء التوقعات او التقدير او في تحليل الحساسية لاغراض السيطرة والتحكم وكما نستدل من عدم معنوية جميع قيم t المتعلقة باختبار المتغيرات التي تضمنها النموذج، فالاهداف غير الوصفية تتطلب فحص الفرضيات واختبارات عملية ومنطقية ومدى معنوية كل من المتغيرات الداخلة في المنوذج وغيرها كما سنرى لاحقا.

وبذلك نحصل على النموذج الاحصائي المبينة معاملاته ومعايير تقييمه في جداول المخرجات. ومن هذه الجداول نجد بان النموذج قد تضمن جميع المتغيرات باستخدام طريقة enter وبالتالي فان المعايير التي جاء بها النموذج هي عالية المعنوية، حيث ان :

$$R = 0.953$$

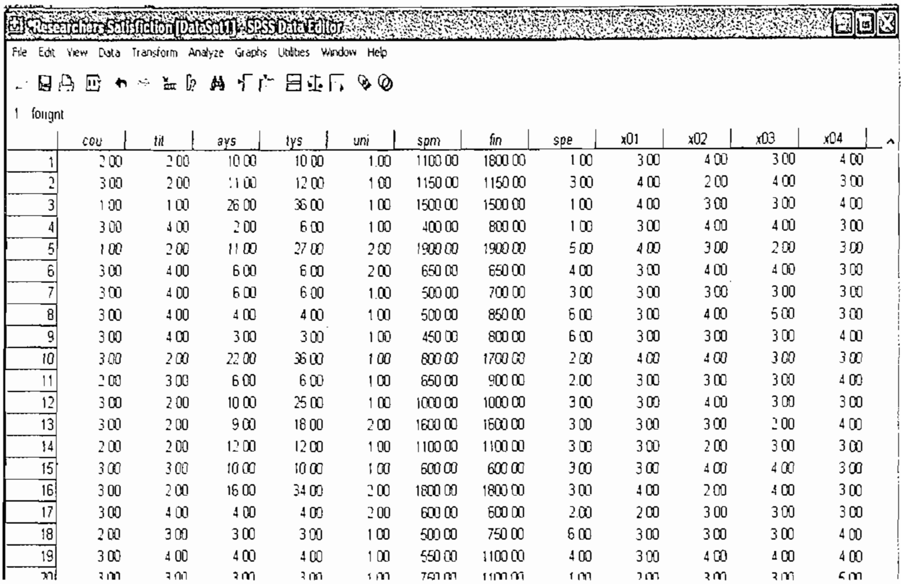
$$R^2 = 0.907$$

$$F = 18.8 \quad , \quad \text{Sig. at } 0.000$$

في الوقت الذي نجد فيه وكما اشرنا، بان كافة معاملات الانحدار بمقياس اختبار t-test هي غير معنوية. بينما سنجد ان النموذج المستوفي للمعايير الاحصائية وللمعاييرالمنطقية (الاشارات) ولفرضيات التوزيع والدقة الذي تم بناؤه في الفصل السادس لايتضمن اكثر من 5 متغيرات معنوية من مجموع 25 متغيرا تم اخضاعها لعملية التحليل.

الشكل البياني رقم (27.5) :

مقطع من ملف معطيات الحالة الدراسية



Researchers Satisfiction [DataSet1] - SPSS Data Editor

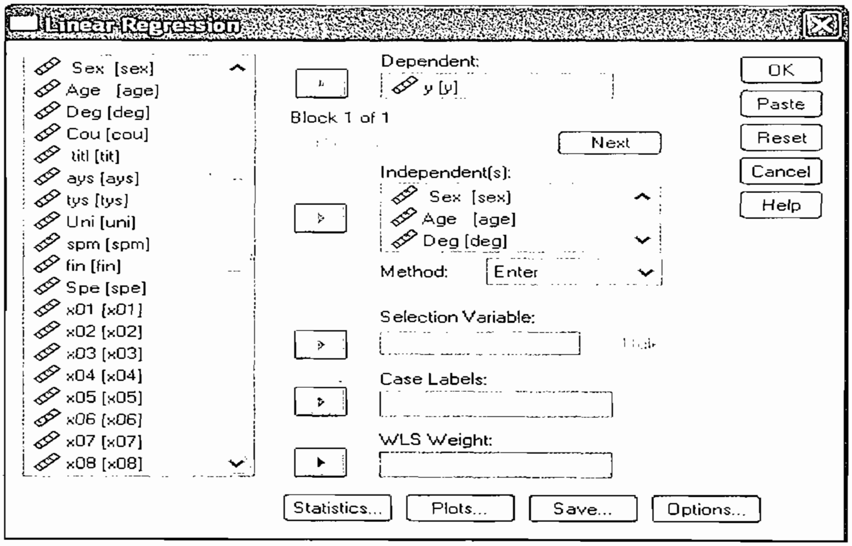
File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : foignt

	cou	tit	ays	tys	uni	spm	fin	spe	x01	x02	x03	x04
1	2.00	2.00	10.00	10.00	1.00	1100.00	1800.00	1.00	3.00	4.00	3.00	4.00
2	3.00	2.00	11.00	12.00	1.00	1150.00	1150.00	3.00	4.00	2.00	4.00	3.00
3	1.00	1.00	26.00	36.00	1.00	1500.00	1500.00	1.00	4.00	3.00	3.00	4.00
4	3.00	4.00	2.00	6.00	1.00	400.00	800.00	1.00	3.00	4.00	4.00	3.00
5	1.00	2.00	11.00	27.00	2.00	1900.00	1900.00	5.00	4.00	3.00	2.00	3.00
6	3.00	4.00	6.00	6.00	2.00	650.00	650.00	4.00	3.00	4.00	4.00	3.00
7	3.00	4.00	6.00	6.00	1.00	500.00	700.00	3.00	3.00	3.00	3.00	3.00
8	3.00	4.00	4.00	4.00	1.00	500.00	850.00	6.00	3.00	4.00	5.00	3.00
9	3.00	4.00	3.00	3.00	1.00	450.00	800.00	6.00	3.00	3.00	3.00	4.00
10	3.00	2.00	22.00	36.00	1.00	800.00	1700.00	2.00	4.00	4.00	3.00	3.00
11	2.00	3.00	6.00	6.00	1.00	650.00	900.00	2.00	3.00	3.00	3.00	4.00
12	3.00	2.00	10.00	25.00	1.00	1000.00	1000.00	3.00	3.00	4.00	3.00	3.00
13	3.00	2.00	9.00	18.00	2.00	1600.00	1600.00	3.00	3.00	3.00	2.00	4.00
14	2.00	2.00	12.00	12.00	1.00	1100.00	1100.00	3.00	3.00	2.00	3.00	3.00
15	3.00	3.00	10.00	10.00	1.00	600.00	600.00	3.00	3.00	4.00	4.00	3.00
16	3.00	2.00	16.00	34.00	2.00	1800.00	1800.00	3.00	4.00	2.00	4.00	3.00
17	3.00	4.00	4.00	4.00	2.00	600.00	600.00	2.00	2.00	3.00	3.00	3.00
18	2.00	3.00	3.00	3.00	1.00	500.00	750.00	6.00	3.00	3.00	3.00	4.00
19	3.00	4.00	4.00	4.00	1.00	550.00	1100.00	4.00	3.00	4.00	4.00	4.00
20	3.00	3.00	3.00	3.00	1.00	750.00	1100.00	1.00	2.00	3.00	3.00	5.00

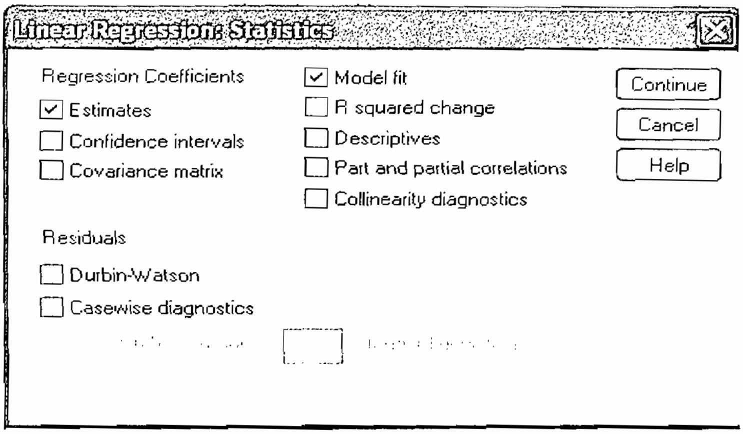
الشكل البياني رقم (28.5)

مربع حوار Linear Regression



الشكل البياني رقم (29.5)

لوحة Linear Regression : Statistics



جداول مخرجات رقم (8.5)

التحليل الوصفي باستخدام تحليل الانحدار

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.953(a)	.907	.859	1.826

a Predictors: (Constant), x09, Spe, x15, x02, x06, Uni, x07, x14, Sex , x04, Cou, x08, x11, x12, x10, ays, x13., x05, Deg, x03, x01, fin, Age , titl, spm

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1567.035	25	62.681	18.804	.000(a)
	Residual	160.005	48	3.333		
	Total	1727.041	73			

a Predictors: (Constant), x09, Spe, x15, x02, x06, Uni, x07, x14, Sex , x04, Cou, x08, x11, x12, x10, ays, x13., x05, Deg, x03, x01, fin, Age , titl, spm
b Dependent Variable: y

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	Constant	10.300	4.581		2.249	.029
	Sex	-.539	.546	-.052	-.987	.329
	Age	.493	.581	.112	.849	.400
	Deg	-1.483	.991	-.153	-1.497	.141
	Cou	-.465	.416	-.073	-1.117	.270
	titl	-.815	.709	-.158	-1.150	.256
	ays	.317	.090	.488	3.526	.001
	Uni	-.199	.385	-.042	-.516	.608
	spm	.001	.001	.148	.715	.478
	fin	.000	.001	-.053	-.339	.736
	Spe	-.213	.152	-.074	-1.403	.167
	x01	.180	.490	.032	.368	.714
	x02	-.023	.327	-.004	-.071	.944
	x03	-.229	.348	-.045	-.657	.514
	x04	-.146	.375	-.022	-.388	.699
	x05	-.434	.376	-.065	-1.153	.255
	x06	.059	.399	.008	.148	.883
	x07	.219	.393	.032	.558	.580
	x08	-.290	.301	-.050	-.963	.340
	x10	.129	.377	.019	.343	.733
	x11	-.143	.305	-.029	-.470	.640
	x12	.000	.382	.000	-.001	.999
	x13.	.193	.345	.036	.559	.579
	x14	.326	.218	.081	1.492	.142
	x15	.323	.357	.052	.904	.370
	x09	-.333	.310	-.065	-1.073	.289

a Dependent Variable: y

5- 4 استخدام تحليل المركبات الاساسية في التحليل الوصفي

Principal Component Analysis for Descriptive Analysis

5- 4- 1 خصائص استخدام المركبات الاساسية في التحليل الوصفي

يعتبر تحليل المركبات (بضم الميم) اهم فصيلة في تحليل العوامل factor analysis، لانها تتميز بامكانية استخدامها مع كافة انواع المعطيات سواء اكانت موزعة توزيعا طبيعيا ام لم تكن كذلك.

وهي طريقة احصائية وصفية تستطيع تصنيف اعداد كبيرة من المتغيرات الى عدد محدود من المركبات اعتمادا على العلاقات التي تربط كل مجموعة من المتغيرات فيما بينهما، وهي بذلك تستطيع تقليص عدد كبير من المتغيرات من دون ان يؤدي ذلك الى فقدان جوهري في نسبة التباين التي يتم تفسيرها .

والمركبات الاساسية Cp هي متغيرات عشوائية غير مترابطة، وكل منها تضم مجموعة متغيرات عشوائية (X1, X2, -----------, Xp) مترابطة وتشترك بأتجاه خطي، بحيث تأخذ الصيغة التالية :

$$C_j = \sum_{i=1}^{P} a_{ij} \quad , \; j = 1,2, \text{--------} , p$$

وان aij هي عبارة عن معاملات عناصر (elements) مصفوفة الموجهات الذاتية الطبيعية (Normalized eigen vectors) لمصفوفة الارتباط للمتغيرات xi's

وبواسطة مصفوفة قيم الموجهات الذاتية (eigen values) نحصل على الحجم النسبي للتغاير او التباين المفسر للبيانات الاحصائية بواسطة كل من المركبات الاساسية التي يتم احتسابها من المتغيرات xi's بأستخدام

التحويـل المتعامـد (orthogonal transformation) ويطلـق علـى المعاملات aij عـادة بتحمـيلات المركبـة (component loadings) وهذه التحميلات تشير الى وزن العلاقة بـين المـتغيرات xi's والمركبـات الاساسية cp' s بشرط اخذ الجذر التربيعي لمصفوفة ارتباط تباينات الشيوع لقيم التبـاين الـذاتي (Morrison, 1967). ويطلـق علـى قـيم الجـذر التربيعي لتباينات الشيوع او التباين المشترك (Communality) ويمكن الرمز له بـ :

$$h_j = aj^2 1 + aj^2 2 + \text{---------} + aj^2 p$$

5-4-2 حالة دراسية رقم C_{5-8} استخدام برنامج SPSS بتوظيف تحليل المركبات في التحليل الوصفي

وسيتم هنا اخضاع المتغيرات البالغ عددها 25 متغيرا التي تم تسميتها في C_{5-7} حالة استخدام الانحدار في التحليل الوصفي للفقرة (5-3) اعلاه.

بعد اخضاع ملف المتغيرات المذكورة والمبين مقطع منه في الشكل البياني رقم (27.5) الى برنامج SPSS ، فان انجاز عملية التحليل تمر بالخطوات التالية :

- استدعاء الامر الفرعي Data Reduction من قائمة Analysis والكبس على خيار Factor فيظهر لنا مربع الحوار Factor Analysis المبين في الشكل البياني رقم (30.5) ليتم فيه استخدام السهم الجانبي لنقل المتغيرات الى تحت عنوان Variables، وان كان هناك متغير محدد مطلوب ظهوره يتم نقله في خانة Selection Variable ،
- الكبس على ايقونة ' Extract فتظهر لوحة خيارات Factor Analysis : Extraction المبينة في الشكل البياني رقم (31.5) يتم فيها

اختيار طريقة Principal Components، والتاشير تحت عنوان Analysis على مصفوفة الارتباط اوغيرها في حالة الرغبة ،والتاشير على Eignvalues over 1 تحت عنوان Extract، والتاشير على Unrotated Factor Solution تحت عنوانDisplay، ثم الكبس على ايقونة Continue للعودة الى مربع الحوار ،

- الكبس على ايقونة Descriptives لتظهر لنا لوحة Factor Analysis : Descriptives المينة في الشكل البياني رقم (32.5)، ويتم فيها التاشير عندInitial Solution تحت عنوان Statistics، والتاشير على المؤشرات المطلوبة مثل المعاملات Coefficients ومستوى المعنوية Level of Significance الواردة تحت عنوان Correlation Matrix، ثم الكبس على ايقونة Continue للعودة الى مربع الحوار مرة اخرى ،
- الكبس على ايقونة Rotation فتظهر لوحة Factor Analysis: Rotation المبينة في الشكل البياني رقم (33.5)، فيتم التاشير على طريقة تحت عنوان ، ثم الكبس على ايقونة للعودة من جديد الى مربع الحوار ،
- الكبس على ايقونة Scores فتظهر لوحة Factor Analysis: Scores المبينة في الشكل البياني رقم (34.5)، ليتم فيها التاشير Display ... Factor Scores، ثم الكبس ايقونة Continue للعودة الى مربع الحوار ،
- الكبس على ايقونة Options فتظهر لوحة Factor Analysis: Options، المبينة في الشكل البياني رقم (35.5)، وفيها يتم التاشير على Sorted by Size ثم الكبس على ايقونة للعودة الى مربع الحوار ،
- الكبس على ايقونة Ok نحصل على مخرجات التحليل، والجداول رقم (9.5) تمثل بعض من المخرجات بقدر تعلق الامر بالتحليل الوصفي .

ومن المخرجات نستدل على ان عدد المركبات المستوفية لمعايير المعنوية وفقا لمعياري قيم الموجهات الذاتية Eigen value والتباين المشترك Communality هي 8 مركبات معنوية والجداول رقم (9.5) توضح بان هذه المركبات الثمانية استطاعت تفسير ما نسبته 70.71 ٪ من مجموع التباين الذاتي Eigen values، وان اعلى نسبة تفسير تحققت هي للمركبتين 1 و 2 اللذان قاما بتفسير ما نسبته 24.83 ٪ و 11.37 ٪ على التوالي .

اما على نطاق المتغيرات، فان اعلى تفسير للتباين المشترك Communalities فهو لكل من المتغيرات التالية :

(0.929) Spm، (0.871) Fin، (0.882) Tit، Age (0.886)، X_{14} (0.697)، اما نسبة تفسير التباين المشترك بواسطة المتغير التابع Y فقد بلغت ما مقداره (0.918) .

وان توصيف هذه المركبات واسلوب اختيار تسميتها هو كما يلي :

المركبة 1 :- ان ابرز مجموعة للمتغيرات المترابطة فيما بينها التي تضمها لمركبة Component تنتمي لمدة الخدمة سواء لمجموع فترة الخدمة Tys (0.926) او الخدمة الاكاديمية Nay (0.92) والى عمر الباحث Age(0.906)، وحيث ان اعلى معامل تحميل Loading جاء به هو متغير Tys ولاهمية المتغير فيمكن اختياره ليكون ممثلا Proxy لباقي المتغيرات في هذه المركبة، واقتراح ان نطلق على المركبة تسمية '**فترة خدمة الباحث**' .

المركبة 2 :- وقد جاءت بمتغيرات اغلبها واهمها مرتبط بموضوع الدخل، فاعلى معاملات تحميل تعود لمتغيرات: الراتب الشهري Spm (0.906)، معدل دخل الاسرة Fin (0.91)، مستوى الرضا عن الراتب الشهري X_{01} (0.805) ولاهمية متغير معدل دخل الاسرة وظهوره باعلى معامل تحميل، فيمكن ان نطلق على المركبة اسم '**الدخل الشهري للاسرة**' ليكون ممثلا لباقي متغيرات المركبة .

المركبة 3 :- وتتمثل اهم المتغيرات في هذه المركبة بمتغيري مستوى الرضا عن توفير التسهيلات الضرورية لاجراء البحوث كالاجهزة والمراجع والدوريات X_{03} (معامله 0.803) وعن توفير العدالة بين الباحثين X_{06} (0.78) سواء في مجال الايفاد وحضور المؤتمرات وما شابه، والتي تعتبر ايضا ضرورة علمية لتنمية قدرات الباحث وتمكينه في العمل البحثي، وكما نلاحظ فان متغير توفير تسهيلات العمل البحثي قد ظهر باعلى معامل ولاهميته فيمكن اختياره واعتباره معبراعن متغيرات المركبة الاخرى وان يقترن اسم المركبة بالمتغير المذكور وتسميتها "**توفير مستلزمات البحوث**".

المركبة 4: -مستوى الرضا عن القوانين والتعليمات التي تنظم العمل البحثي في الجامعات X_{12} ومعامله (0.661) و متغير مستوى الرضا عن المحفزات المالية والمعنوية X_{10} (0.621). ومن ذلك يمكن الاستدلال على اهمية شمول القوانين والتعليمات على المحفزات المالية والمعنوية او القيام بتفعيلها في حالة كانت مشمولة، وعليه يمكن اقتراح تسمية المركبة "**القوانين ومحفزات البحوث**".

المركبة 5 :- والمعاملات المعنوية في هذه المركبة تعود لمتغيري مستوى الرضا عن الاهتمام بالكفاءات العلمية البحثية X_{08} (0.725) ومستوى الرضا عن الحصول على حقوق الترقية العلمية X_{05} (0.718)، وهذه التحميلات تقترح تسمية المركبة "**رعاية الكفاءات وحقوق ترقيتهم**".

المركبة 6 :- واعلى المعاملات في المركبة هي لمتغيري كل من مستوى الرضا عن الموازنة بين عدد المحاضرات المطلوب اداءها من الكادر البحثي والمهام البحثية التي يقوم بها X_{07} (-0.62)، وقد جاء المتغير باشارة سالبة والتي تدل من خلال ترميز مستوى الرضا من اعلى الى الادنى على ان هذه الموازنة قد تكون غير متحققة للالقاب العلمية الدنيا حيث يزداد عليها عبئ المحاضرات الذي قد يصل لاكثر من ضعف عما هو مطلوب من لقب استاذ

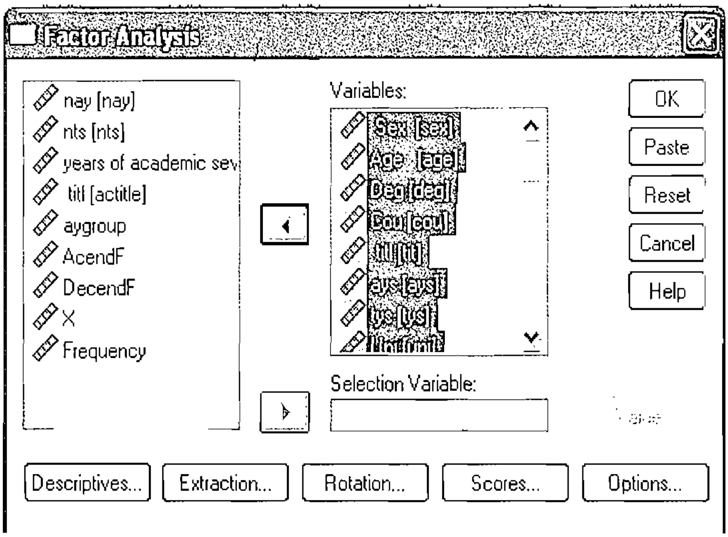
مثلا في اغلب الجامعات، اما المتغير الاخر فهو الاختصاص Spe (0.67). ولكون الاخير ظهر باعلى معامل تحميل، عليه يمكن اقتراح تسميته بمركبة **"الاختصاص وعدد المحاضرات"**.

المركبة 7 :- والمتغيرات المعنوية التي جاءت في هذه المركبة هي لكل من الجنس Sex وجاء باشارة سالبه (-0.733) ومتغير مستوى الرضا عن جدية الجامعات بتشجيع البحوث X_{11} (0.643)، ووفقا لقيم الترميز الذي ياخذ الرقم 2 بالنسبة للاناث، يستدل بان الباحثين من الاناث يرتفع لديهم شعور عدم القناعة من جدية الجامعات بقيامهن بالعمل البحثي. وان تلازم فحوى كلا المتغيرين تساعد على اقتراح تسمية المركبة **"الجنس"**.

المركبة 8 :- وهي المركبة التي خصت متغير مستوى الرضا عن مدى جدية الجهات المستفيدة في تطبيقها لنتائج البحوث المنشورة X_{14} ومعامل حجمه 0.83، والخيار هو تسمية المركبة **"مدى الجدية بتطبيق نتائج البحوث"**. ليكون ممثلا لباقي متغيرات المركبة

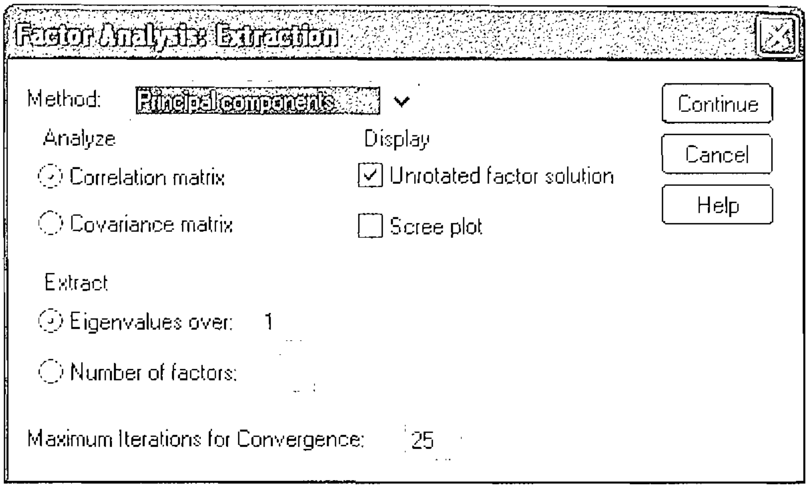
الشكل البياني رقم (30.5)

مربع الحوار Factor Analysis



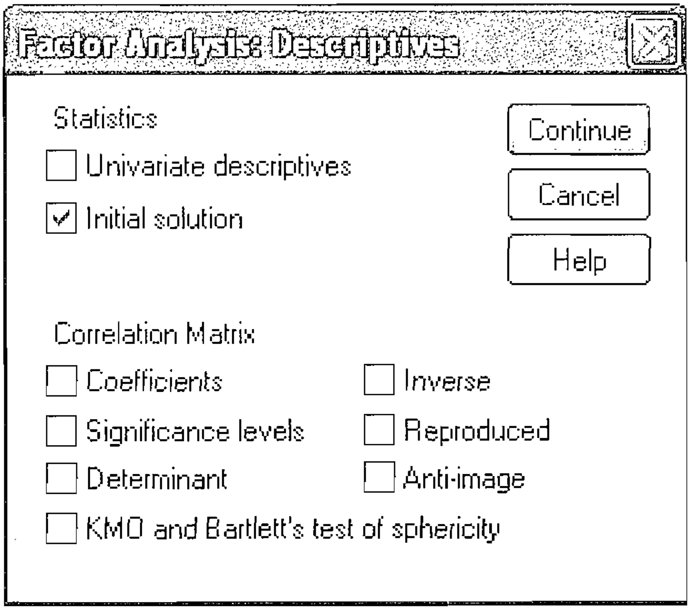
الشكل البياني رقم (31.5)

لوحة خيارات Factor Analysis : Extraction



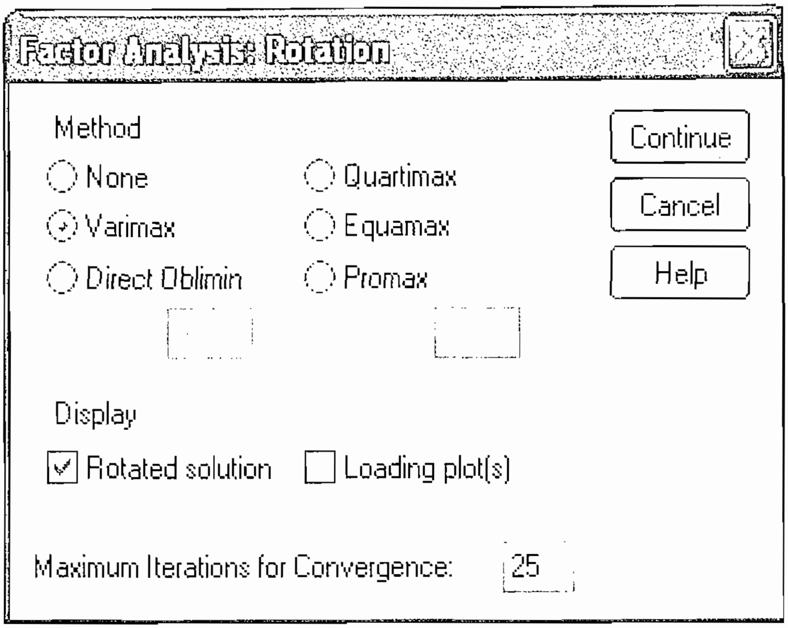
الشكل البياني رقم (32.5)

لوحة Factor Analysis : Descriptive



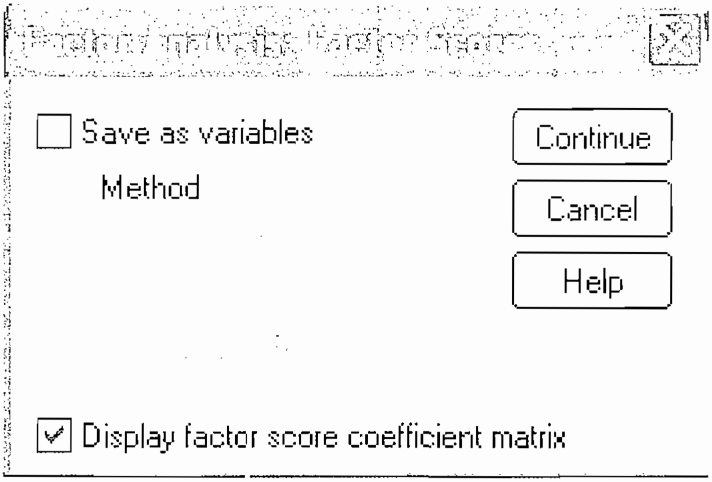
الشكل البياني رقم (33.5)

Factor Analysis: Rotation لوحة



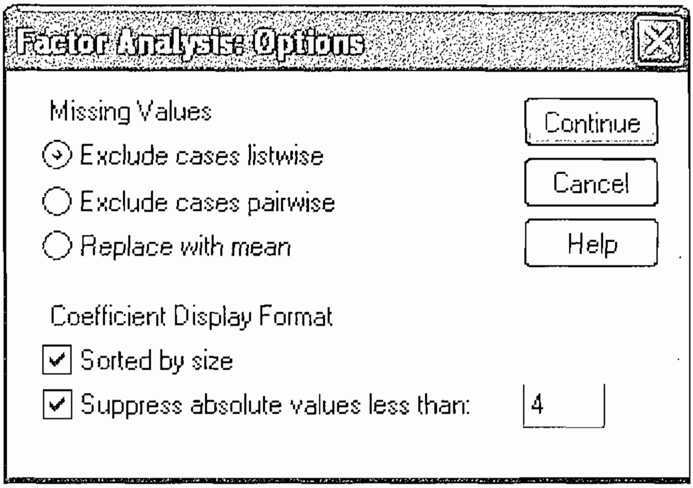
الشكل البياني رقم (34.5)

Factor Analysis: Scores لوحة



الشكل البياني رقم (35.5)

فتظهر لوحة Factor Analysis: Options



جداول رقم (9.5) بعض من مخرجات تحليل المركبات

Communalities

Variable	Extraction	Variable	Extraction
Sex	.651	x03	.738
Age	.886	x04	.577
Deg	.655	x05	.662
Cou	.595	x06	.646
titl	.882	x07	.628
ays	.867	x08	.625
tys	.901	x09	.666
Uni	.653	x10	.535
spm	.929	x11	.719
fin	.871	x12	.561
Spe	.541	x13.	.697
x01	.771	x14	.697
x02	.551	x15	.682
y	.918	nay	.676

Extraction Method: Principal Component Analysis

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.704	24.831	24.831	6.704	24.831	24.831
2	3.069	11.368	36.199	3.069	11.368	36.199
3	2.353	8.715	44.914	2.353	8.715	44.914
4	1.771	6.560	51.474	1.771	6.560	51.474
5	1.543	5.714	57.188	1.543	5.714	57.188
6	1.346	4.984	62.172	1.346	4.984	62.172
7	1.244	4.608	66.780	1.244	4.608	66.780
8	1.072	3.971	70.751	1.072	3.971	70.751

Rotated Component Matrix(a)

	Component							
	1	2	3	4	5	6	7	8
ays	.939							
nay	.933							
Age	.929							
tys	.927							
y	.894							
titl	-.848	-.320						
Deg	-.504	-.418					-.313	
spm		.914						
fin		.903						
x01		.789						
Uni		.640		-.371				
Cou	-.443	-.546						
x09		.475			.303		.390	
x03			.799					
x06			-.774					
x04			-.492			.483		
x12				.667				
x10				.636				
x13.				.551			.393	
x15				.509	-.483			
x08					.748			
x05					.693			
Spe						.658		
x07	.338			-.326		-.621		
x02	-.376			.303		-.402		
x14							.807	
Sex							-.302	-.721
x11	.306						-.313	.641

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 8 iterations.

الفصل السادس

أساليب بناء النماذج الاحصائية

لاغراض التوقعات، التقدير، والسيطرة والتحكم

Methods of Statistical Model Building For Forecasting, Estimation and Controlling

1-6 تحليل الانحدار Regression Analysis

يبحث الانحدار في العلاقة بين المتغيرات من خلال بناء معادلة تستخدم للتفسير او للتقدير اوالتوقع او التنبوء بقيمة المتغير التابع Y بدلالة متغير او متغيرات مستقلة Xi، ويمكن اجمال اهداف تحليل الانحدار بما يلي :

- تحديد العلاقة بين المتغير التابع Y ومتغير مستقل X او اكثر
- التنبوء بالمتغير التابع بدلالة متغير مستقل او اكثر باستخدام العلاقة التقديرية
- الاستدلال حول المجتمع ووصفه من خلال المعادلة التقديرية
- اختبار الفروق بين خطي الانحدار التقديري والحقيقي
- كأداة للسيطرة والتحكم باتجاه دالة معينة وحجمها

2-6 تحليل الانحدار الخطي البسيط

Simple Linear Regression Analysis

6 -2 -1 معادلة الانحدار الخطي البسيط

الانحدار الخطي البسيط Simple Linear Regression يعني البحث في العلاقة بين متغيرين فقط هما المتغير التابع Y والمتغير المستقل X، وان شكل معادلة العلاقة للمجتمع هي :

$$Y = \alpha + \beta X$$

حيث ان :

Y يدعى بالمتغير التابع او المعتمد Dependent Variable

α يدعى بالمعامل الثابت Constant Coefficient ويصبح مساويا لقيمة Y عندما قيمة X تساوي صفر ،

β يدعى بميل الانحدار، ويمثل مقدار التغير في Y عند زيادة قيمة المتغير المستقل بمقدار 1

X يدعى بالمتغير المستقل،

ويستعاض عن الحرف Y بـ y عندما تكون معطيات القيم الحقيقية تعود لعينة في بناء المعادلة. وعند بناء المعادلة التقديرية التي تعتمد على معطيات عينة يصبح شكل المعادلة كالاتي :

$$\hat{y} = a + bx$$

وحيث من غير المتوقع ان تقع النقاط تماما على خط الانحدار، فان العلاقة الخطية التامة يتم تعديلها لكي تضم متغير خطأ عشوائي يرمز له بـ ε_I (يكون الرمز e_i في المعادلة التقديرية) يمثل انحراف القيم التقديرية عن القيم الحقيقية y ويمكن التعبير عن هذه العلاقة بالصيغة التالية :

$$\hat{y} = a + bx + e_i$$

ويفترض في المتغير العشوائي ε_I انه موزع طبيعيا، وقيمته المتوقعه (وسطه) مساويا للصفر، وتباينه ثابت، وان هذه الاخطاء غير مترابطة ببعضها، وهو ما يطلق عليه بـ Multicollinearity، وانه مستقل، اي ان عملية السحب العشوائي لاتعتمد على بعضها لكي لايحصل ترابط بين قيم المتغير المستقل وقيم متغير الخطأ العشوائي، وسيتم التطرق بتفصيل وافي عن هذه الفرضيات Assumptions لاحقا من هذا الفقرة .

6- 2- 2 طريقة المربعات الصغرى وخواصها

Least Square Method and Propreties

يتم استخدام طريقة المربعات الصغرى في تقدير ميل الانحدارغير المعلوم وهي الطريقة التي تقوم بتقليل مجموع مربعات انحرافات القيم الحقيقية y عن القيم التقديرية $\hat{y}$.

(1) خطوات طريقة المربعات الصغرى

بما ان قيم المعطيات x_i , y_i يكون انتشارها حول خط الانحدار، فان انحراف كل قيمة حقيقية لـ y_i عن القيمة التقديرية $\hat{y}_i$ يكون بمقدار e_i ، اي :

$$e_i = y_i - \hat{y}_i$$

وعليه :

- يتم تربيع مجموع قيم طرفي المعادلة السابقة وتعويض y_i، فنحصل على :

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

$$= \sum (y_i - \hat{y}_i)^2$$

- يجري تقليل مجموع مربعات e_i باستخدام طريقة المربعات الصغرى من خلال تفاضل ei لكل من a و b بمساواتهم للصفر، اي :

$$\frac{\partial e}{\partial a} = -2\sum (y_i - a - bx_i) = 0$$

$$\frac{\partial e}{\partial b} = -2\sum x_i (y_i - a - bx_i) = 0$$

فيكون لدينا : $ny_i - na + b\sum x_i = 0$

$$\sum x_i y_i = a + \sum x_i + b\sum x_i^2$$

- ومنه يتم تقدير قيم معاملات a و b كالاتي :

$$b=\frac{\Sigma x_i y_i \ \frac{\Sigma x_i \Sigma y_i}{n}}{\Sigma x_i^{2}-\frac{\Sigma x_i^{2}}{n}}$$

$$=\frac{n\Sigma x_i y_i-\Sigma x_i \Sigma y_i}{n\Sigma x_i^{2}-(\Sigma x_i)^{2}}$$

- وبقسمة البسط والمقام على n نحصل على :

$$b=\frac{\frac{\Sigma x_i y_i}{n}-(\frac{\Sigma x_i}{n})(\frac{\Sigma y_i}{n})}{\frac{\Sigma x_i^{2}}{n}-(\frac{\Sigma x_i}{n})^{2}}$$

$$=\frac{\frac{\Sigma x_i y_i}{n}-\overline{x}\overline{y}}{\frac{\Sigma x_i^{2}}{n}-(\frac{\Sigma x_i}{n})^{2}}$$

- فيكون لدينا : $\overline{y}=0$ و $\overline{x}=0$ من البسط والمقام، وحيث ان n وبحذف

$$b=\frac{\Sigma x_i y_i}{\Sigma x_i^{2}}$$

وحيث ان : $\Sigma y_i = na + b\Sigma x_i$

فان : $a=\frac{\Sigma y_i - b\Sigma x_i}{n}=\frac{\Sigma y_i}{n}-b\frac{\Sigma x_i}{n}$

فنحصل على : $a=\overline{y}-b\overline{x}$

(2) خواص مقدرات طريقة المربعات الصغرى

▫ خاصية عدم التحيز، وتعني ان القيمة المتوقعة لكل من a و b مساوية للقيم الحقيقية، أي :

$$E(b) = b$$

$$E(a) = a$$

حيث ان :
$$E(b) = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}$$

$$= a \frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} + b \frac{\sum (x_i - \bar{x}) x_i}{\sum (x_i - \bar{x})^2} + e_i \frac{\sum (x_i - \bar{x}) e_i}{\sum (x_i - \bar{x})^2}$$

وحيث ان :
$$\sum (x_i - \bar{x}) = 0$$

وان :
$$\sum (x_i - \bar{x}) x = \sum (x_i - \bar{x})^2$$

نحصل على :
$$E(b) = b + \frac{\sum (x_i - \bar{x}) e_i}{\sum (x_i - \bar{x})^2}$$

وبما ان :
$$E(e_i) = 0$$

فان :
$$E(b) = b + \frac{\sum (x_i - \bar{x}) \; E(e_i)}{\sum (x_i - \bar{x})^2} = b$$

وبذلك تكون قيمة b غير متحيزة عن القيمة الحقيقية. ولاثبات خاصية عدم التحيز لـ a لدينا :

$$a = \bar{y} - b\bar{x}$$

وحيث ان :
$$\bar{y} = a + b\bar{x} + \bar{e}$$

وعلى افتراض ان :
$$E(\bar{e}) = \frac{1}{n} E\left(\sum_{i=1}^{a} e_i \right) = 0$$

عليه فان :

$$E(a) = a + b - E(b)\bar{x} - \bar{e}$$

$$= a + xE(b - b)x + E(\bar{e})$$

وان :

$$E(\bar{e}) = 0$$

$$E(b - b) = 0$$

فان :

$$E(a) = a$$

- **خاصية اقل تباين ممكن،** ويقصد بذلك ان يكون تباين كل من a و b أقل مايمكن لاكتساب خاصية التقدير الجيد. ويمكن اثبات ذلك كالاتي :

- **بالنسبة لـ b :**

$$E(b) - b = \frac{\sum(x_i - \bar{x})e_i}{\sum(x_i - \bar{x})^2}$$

$$[E(b) - b]^2 = \left[\frac{\sum(x_i - \bar{x})e_i}{\sum(x_i - \bar{x})^2}\right]^2$$

وباستخدام الفروض :

$$E(x_i e_i) = 0$$

$$E(e_i^2) = s_e^2$$

$$E(e_i e_j) = 0, i \neq j$$

فأن :

$$var(Eb) = E\left[\frac{\sum(x_i - \bar{x})e_i}{\sum(x_i - \bar{x})^2}\right]^2$$

$$= \frac{s_e^2}{\sum(x_i - \bar{x})^2}$$

- **اما بالنسبة الى a، لدينا :**

$$E(a) = a + [b - E(b)]\bar{x} + \bar{e}$$

حيث ان :

$$\bar{e} = \frac{1}{n}\sum \bar{e}_i$$

$$\operatorname{var}(\bar{e}) = E\left[\frac{1}{n}\sum \bar{e}_i\right]^2$$

فان :

$$E\left[\frac{1}{n}\sum \bar{e}_i\right]^2 = \frac{s_e^2}{n}$$

وباستخدام الفروض :

$$E(e_i^2) = s_e^2$$

$$E(e_i e_j) = 0, i \neq j$$

وحيث ان :

$$E(a) - a = \bar{e} - [E(b) - b]\bar{x}$$

اذن :

$$[E(a) - a]^2 = \bar{e}^2 + [E(b) - b]^2 \bar{x}^2 - 2\bar{e}[E(b) - b]\bar{x}$$

وبذلك فان :

$$E(a - a)^2 = \operatorname{var} E(a)$$

$$= E(\bar{e}^2) + \bar{x}^2 E(b - b)^2$$

$$\operatorname{var}(b) = E(b - b)^2$$

وسحيث ان :

$$\operatorname{var}(b) = \frac{s_e^2}{\sum (x_i - \bar{x})^2}$$

فان :

$$\operatorname{var}(b) = \frac{s_e^2}{n}\left[1 + \frac{n\bar{x}}{\sum (x_i - \bar{x})^2}\right]$$

$$= \frac{s_e^2}{n}\left[\frac{\sum (x_i - \bar{x}) - n\bar{x}^2}{\sum (x_i - \bar{x})^2}\right]$$

حيث ان :

$$\sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$$

وعليه يكون لدينا :

$$\operatorname{var}(a) = \frac{s_e^2 \sum x_i^2}{n\sum (x_i - \bar{x})^2}$$

6- 2- 3 فرضيات نموذج الانحدار الخطي البسيط

ان بناء معادلة (او نموذج) الانحدار عادة ما تعتمد على تحليل مشاهدات عينة مسحوبة عشوائيا من مجتمع احصائي، ويتم الاعتماد على نتائج تحليل العينة لتعميمها على المجتمع، وعليه فان عملية التحليل لابد ان تضمن التمثيل التقريبي للمجتمع المسحوبة منه. وحيث انه من غير المتوقع ان تكون العينة ممثله تماما لخصائص المجتمع، لذلك فان بناء نموذج الانحدار الخطي يجب ان يكون مستوفيا لعدد من الفرضيات التي يمكن اجمالها بما يلي :

(1) الفرض الاول، يتعلق بقيم المتغير المستقل x على انها مستقلة، والافتراض هو ان معطيات المتغير قادرة على اظهار تاثيرها في تغير قيم المتغير التابع y، بحيث تكون قيمة واحدة على الاقل من قيم المتغير المستقل مختلفة عن بقية القيم، ويمكن التعبير عن هذا الفرض بالصيغة التالية :

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \neq 0$$

فعندما تكون هناك اخطاء في قياس المتغيرات المستقلة سيؤدي الامر الى خرق فرض استقلالية المتغيرات المستقلة مما يؤدي الى ان تكون تقديرات المعالم متحيزة وغير متسقة، فتكون b متحيزة الى ادنى، بينما a تكون متحيزة الى الاعلى، وليس هناك اختبار محدد للكشف عن وجود مثل هذه الاخطاء ولكن يمكن الاستدلال عليها من الطريقة التي جمعت بها المعطيات .

ويمكن تصحيح مثل هذه الاخطاء بايجاد انحدار x على y. مع الاشارة الى ان اخطاء القياس في المتغير التابع y لاتؤدي الى تحيز في التقديرات لانها تدخل في الخطأ العشوائي e_i .

(2) الفرض الثاني: هو ان الخطأ العشوائي e_i يتبع التوزيع الطبيعي، وكنتيجة فان المتغير التابعy وتوزيع المعاينة لمعالم الانحدار تتبع ايضا التوزيع

الطبيعي، بحيث يمكن اجراء الاختبارات لمعنوية هذه المعالم، وعادة ما يشار الى هذا التوزيع بـ :

$$e_i \sim N(0, S_e^2)$$

(3) الفرض الثالث: هو ان القيمة المتوقعة للخطأ العشوائي (اي وسطه) مساويا للصفر، اي :

$$E(e_i) = 0$$

وبسبب هذا الفرض فان المعادلة $y = a + bx$ تعطي متوسط قيمة y، حيث انه يفترض بان x ثابتة في حين y في المعادلة $y = a + bx + e_i$ تتغير فوق او تحت وسطها مع زيادة او نقصان e_i عن الصفر .

(4) الفرض الرابع، وهو ان تباين حد الخطأ العشوائي ثابت في كل فترة لكافة قيم X، اي :

$$E(e_i)^2 = s_e^2$$

ويكفل هذا الفرض ان كل مشاهدة يمكن الاعتماد عليها بنفس القدر بحيث تكون تقديرات معاملات الانحدار كفوءة وتكون اختبارات الفروض الخاصة بها غير متحيزه، اي :

$$e_i \sim N(0, S_e^2)$$

(5) الفرض الخامس، هو ان القيمة التي ياخذها الخطأ العشوائي في فترة ما تكون غير مرتبطة او متعلقة بقيمته في فترة اخرى، اي :

$$\text{for} \quad i \neq j \quad i,j = 1, 2, \ldots n \; E(e_i, e_j) = 0$$

وهذا يكفل بان تكون القيمة المتوسطة للمتغير التابع y تعتمد على X فقط وليس على e_i، وهو امر مطلوب للحصول على تقديرات كفوءة لمعاملات الانحدار واختبارات غير متحيزة لمعنوياتها .

ولاجل التاكد من استيفاء نموذج الانحدار الخطي للفرضيات اعلاه، يجري تقييم النموذج في ضوء كل من هذه الفرضيات، فاذا كانت العلاقة بين y و x خطية فيفترض ان يكون ميل خط الانحدار β الذي يصف هذه العلاقة اما موجبة او سالبة، وان قيمته مساوية للصفر، فاذا كانت $\beta = 0$ فهذا يعني بان كفاءة معادلة الانحدار في التنبوء او التقدير هي محدودة، عندها تجري محاولة بناء نموذج غير خطي للتاكد من امكانية تحسين كفاءة النموذج رغم ان العلاقة كانت خطية، وفي مثل هذه الحالة فان اختبار فرضية : $H_0 : \beta = 0$ سوف لايتم رفضها، وبذلك نفترض وجود خطأ من النوع الثاني II وان ذلك يعود اما لكون العلاقة الخطية بين y و x هي ليست قوية او انها غير خطية، وتتم عملية الاختبار باستخدام اما الاحصاءه F او الاحصاءه t.

6- 2- 4 اختبار فرضيات نموذج الانحدار الخطي البسيط

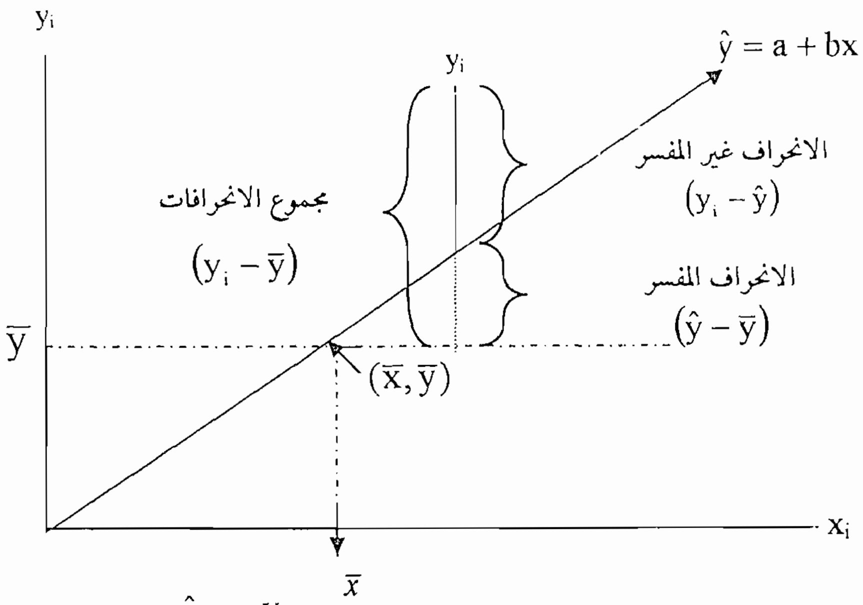
(1) اختبار فرضية $H_0 : \beta = 0$ باستخدام اختبار F

عند التمعن بالشكل البياني (1.6) فان قياس مسافة اية نقطة لـ y_i عن الخط $\bar{y}$ يدعى بمجموع الانحراف ويرمز له بـ $(y_i - \bar{y})$، وان المسافة من خط الانحدار التقديري $\hat{y}$ الى الخط $\bar{y}$ نطلق عليها بالانحراف المفسر ويرمز له بـ $(\hat{y} - \bar{y})$، بينما المسافة من اي نقطة حقيقية عن خط الانحدار $(y_i - \hat{y})$ تدعى بالانحراف غير المفسر، وبذلك فان مجموع الانحراف لاية قيمة y_i تساويس مجموع الانحراف المفسر زائدا الانحراف غير المفسر، اي :

$$(y_i - \bar{y}) = (\hat{y} - \bar{y}) + (y_i - \hat{y})$$

الشكل البياني رقم (1.6)

يوضح مكونات انحرافات نموذج الانحدار



وعند استخراج الانحرافات لكافة قيم y_i، $\hat{y}$ وتربيع كميات الطرفين

نحصل على : $\sum(y_i - \bar{y})^2 = \sum(\hat{y} - \bar{y})^2 + \sum(y_i - \hat{y})^2$

اي الحصول على مقاييس التباين وكالاتي :

- **اجمالي مجموع مربعات التباين** ونرمز له بـSST كمقياس لتشتت القيم الحقيقية حول وسطها الحسابي $\bar{y}$، اي :

$$SST = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$$

- **مجموع مربعات التباين المفسر** ونرمز له بـ SSR وهو مجموع الانحراف المفسر بواسطة علاقة الانحدار الخطية بين قيم المتغير التابع والمتغير المستقل، اي :

$$SSR = b^2\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]$$

- مجموع مربعات التباين غير المفسر ونرمز له بـ SSE وهو عبارة عن مقياس التشتت للقيم الحقيقية حول خط الانحدار، وتعرف بمجموع مربعات البواقي والتي يتم تقليلها باستخدام طريقة المربعات الصغرى، اي:

$$SSE = SST - SSR$$

ويمكن تبويب هذه العلاقة بجدول تحليل التباين الذي ياخذ الشكل التالي:

مصدر التباين	SS	درجات الحرية(d.f)	MS	F
الانحدار الخطي	SSR	1	SSR/1	$F = \frac{MSR}{MSE}$
البواقي	SSE	n-2	SSE/(n-2)	
المجموع	SST	n-1	----	----

فاذا كان نموذج الانحدار معنوي في وصفه للعلاقة بين x و y فان التباين المفسر سيساهم بنسبة كبيرة في تفسير مجموع مربعات التباين، وان مقياس ذلك هو ما يدعى بمعامل التحديد Determination Coefficient ويرمز له بـ r^2 وصيغة ايجاده هي :

$$r^2 = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

$$= \frac{b^2\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]}{\sum y_i^2 - \frac{(\sum y_i)^2}{n}} = \frac{SSR}{SST}$$

وهذا يعني بانه كلما ازدادت قيمة r^2 تقترب القيم الحقيقية في مطابقة القيم التقديرية المستخرجة بواسطة معادلة الانحدار، بكلمة اخرى تقل مسافات ابتعاد y_i عن خط الانحدار.

(2) الخطأ المعياري لميل الانحدار

Standard Error of Regression Slop

ان ميل انحدار العينة، b سيتراوح حول القيم الحقيقية للمجتمع الاحصائي β، ولقياس مقدار انحراف هذا الميل الذي هو b عن ميل المجتمع β نلجأ الى مقياس الخطأ المعياري لميل الانحدار ويرمز له بـ s_b وصيغته :

$$s_b = \frac{s_e}{\sqrt{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}}$$

(3) اختبار فرضية ان المعطيات موزعة طبيعيا

Normal Assumption Testing

ويمكن اعتماد احصاءة الاختبار t لهذا الغرض، وذلك لاختبار فرضية ان نموذج الانحدار معتمد في بناءه على معطيات موزعة توزيعا طبيعيا، وان صيغة الاختبار هي :

$$t = \frac{b - \beta}{s_b}$$

مع درجات حرية $v = n - 2$

والحالات التي يمكن ان تكون عليها الفرضية عند مستوى معنوية α هي :

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

فيتم رفض H_0 اذا كانت $|t| \neq t_{\alpha/2}, \nu$

$H_0 : \beta \leq 0$

$H_1 : \beta \lhd 0$

ويتم رفض H_0 اذا كانت $|t| \rhd t_{\alpha/2}, \nu$

$H_0 : \beta \geq 0$

$H_1 : \beta \lhd 0$

ويتم رفض H_0 اذا كانت $|t| \lhd t_{\alpha/2}, \nu$

حيث ان القيمة صفر تعني عدم وجود فروق بين المتغيرات تحت الاختبار .

(4) مدى الثقة لميل انحدار المجتمع β

وتدلنا عن المدى الذي يقع فيه الميل الحقيقي للمجتمع β عند معامل ثقة معينة، ويعبر عن هذا المدى بالصيغة التالية :

$$b - t_{(1-\alpha/2)}, \nu . S_b \leq \beta \leq b + t_{(1-\alpha/2)}, \nu . S_b$$

(5) الخطأ المعياري لقيمة متوسط المتغير التابع $S_{\bar{y}}$

ويقصد به قياس انحراف القيمة التقديرية لـ $\hat{y}$ عن قيمة متوسط المجتمع الحقيقي $\bar{y}$ وذلك باستخدام الخطأ المعياري التقديري ويرمز له بـ $S_{\bar{y}}$ وصيغته هي :

$$s_{\bar{y}} = s_e \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

حيث ان x تشير الى القيمة المطلوب تعويضها للمتغير المستقل .

(6) مدى الثقة لوسط المجتمع $\bar{Y}$

ويتم حساب المدى الذي سيقع ضمنه وسط المجتمع $\bar{Y}$ بموجب الصيغة التالية

$$6 \quad \bar{y} - t_{(1-\alpha/2)}, v.s_{\bar{y}} \leq \bar{Y} \leq \bar{y} + t_{(1-\alpha/2)}, v.s_{\bar{y}}$$

(7) معامل الثقة للمعامل الثابت α

وتاخذ صيغة حساب معامل الثقة للمعامل الثابت الشكل التالي :

$$a - t_{(1-\alpha/2)}, v.s_{\bar{y}} \leq \alpha \leq a + t_{(1-\alpha/2)}, v.s_{\bar{y}}$$

(8) أستخدام نموذج الانحدار الخطي البسيط في التنبوء

عقب تقييم نموذج الانحدار والتاكد من استيفاءه للفرضيات والمعايير الاحصائية، يصبح بالامكان استخدامه لاغراض التنبوء وذلك بايجاد قيم المتغير التابع y بتغيير قيم المتغير المستقل x. وتتمثل عملية التنبوء بتعويض القيم المطلوبة بـ x للحصول على قيم y . ويكون مدى الثقة لقيمة التنبوء $\hat{y}$ مشابهه لتلك المتعلقة بمدى الثقة لمتوسط المجتمع $\bar{Y}$ باستثناء احتساب قيمة اضافية وهي الخطأ المعياري التقديري $s_{\frac{y}{x}}$ وصيغتها هي :

$$s_{\frac{y}{x}} = s_e \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

وبذلك تكون صيغة مدى الثقة للقيمة الحقيقية المتنبأ بها على الشكل التالي:

$$\hat{y} - t_{(1-\alpha/2)}, vs_{y/x} \leq \hat{Y} \leq \hat{y} + t_{(1-\alpha/2)}, vs_{y/x}$$

مثال (1-6) :

المعطيات في الجدول رقم (1.6) تمثل كمية انتاج الشعير (بالاف الاطنان)، y_i والمساحة المزروعة (بالاف الهكتارات)، x_i مصنفة حسب عينة البلديات المشمولة لدولة ما. والمطلوب :

- ايجاد معادلة الانحدار الخطي التقديرية .
- استخدام الاحصاءة F لاختبار معامل الانحدار عند مستوى معنوية $\alpha = 0.05$
- ايجاد الخطأ المعياري لميل انحدار النموذج، S_b
- اختبار فرضية من ان ميل انحدار المجتمع $\beta \leq 0.5$ وعند مستوى معنوية $\alpha = 0.05$
- ايجاد مدى الثقة لميل انحدار المجتمع β عند ثقة مقدارها 90 ٪
- ايجاد الخطأ المعياري لقيمة متوسط المتغير التابع $s_{\bar{y}}$ عند $x = 75.5$
- ايجاد مدى الثقة لمتوسط المجتمع $\bar{Y}$ والمعامل الثابت α عند ثقة مقدارها 90 ٪ .

جدول رقم (1.6)

كمية انتاج الشعير (بالاف الاطنان)، y_i والمساحة المزروعة (بالاف الهكتارات) x_i لعينة من بلديات احدى الدول

البلدية	انتاج الشعير، y (الاف الاطنان)	المساحة المزروعة، x (الاف الهكتارات)
1	133.3	56.5
2	606.5	175.6
3	375.5	85.5
4	277.0	75.5
5	336.5	111.3
6	255.8	25.4
7	241.4	17.8
8	130.2	48.5
9	62.8	24.1
10	88.3	2.3
11	13.8	1.1
12	22.6	2.3
13	103.7	4.2

الحل لـ (1.6) :

◄ ايجاد معادلة الانحدار الخطي البسيط التقديرية

- لدينا :

$$\sum y_i = 2647.4$$
$$\sum x_i = 629.8$$
$$\sum x_i^2 = 63299.08$$
$$\sum xy = 223719.61$$
$$\left(\sum x\right)^2 = 396648.04$$

- نجد قيم كل من a , b

$$b = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^{\ 2} - (\Sigma x_i)^2}$$

$$= \frac{13(223719.61) - (629.8)(2647.4)}{13(63299.08) - 396648.04} = \frac{1241022.4}{426240} = 0.291$$

$$s\,a = \frac{\sum y - b\sum x}{n}$$

$$= \frac{2647.4 - (0.291)(629.8)}{13} = 189.539$$

- وبذلك تكون معادلة الانحدار الخطي البسيطة التقديرية هي :

$$\hat{y} = 189.539 + 0.291x$$

◄ استخدام الاحصاءة F لاختبار معامل الانحدار عند مستوى معنوية $\alpha = 0.05$

- الفرضية

$$H_0 : \beta = 0$$
$$H_1 : \beta \neq 0$$

من الجدول نجد ان قيمة $F_{0.025}, 1, 11 = 6.724$

- نجد اجمالي مجموع مربعات التباين SST

$$SST = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$$

$$= (133.3)^2 + (606.5)^2 + ...+ (103.7)^2 - \frac{(2647.7)^2}{n}$$

$$= 880414.43 - 539255.01 = 341159.41$$

- نجد مجموع مربعات التباين المفسر SSR

$$SSR = b^2\left[\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right]$$

$$= (0.291)^2\left[(56.5)^2 + (175.6)^2 + ...+ (4.2)^2 - \frac{(629.8)^2}{n}\right]$$

$$= 63299.08 - 30511.387 = 32787.7$$

- مجموع مربعات التباين غير المفسر SSE

$$SSE = SST - SSR$$

$$= 341159.42 - 32787.7 = 308371.71$$

- وبتبويب النتائج اعلاه في جدول تحليل التباين يكون لدينا :

F	MS	درجات الحرية (d.f.)	SS	مصدر التباين
$\frac{32787.7}{28033.79} = 1.17$	32787.7	1	32787.7	الانحدار الخطي SSR
	28033.79	11	308371.71	البواقي SSE
------	------	12	341159.41	المجموع SST

وحيث ان قيمة F المحتسبة أقل من قيمة $F_{0.025},1,11 = 6.724$ الجدولية، عليه نقبل فرضية العدم H_0 مما يستدل على خطية معادلة الانحدار، ولكون β مقارب او مساوي للصفر، فهذا يعني بان المعادلة محدودة الكفاءة في بناء

التنبوءات، مما يستوجب محاولة بناء نموذج غير خطي. وللتحقق من مدى معنوية النموذج في وصفه للعلاقة بين y و x نقوم باستخدام معامل التحديد r^2 وكالاتي :

$$r^2 = \frac{SSR}{SST} = \frac{32787.7}{341159.41} = 0.096$$

ومنه نستدل على صعف التباين المفسر بواسطة التموذج، مما يدل على ابتعاد قيم y_i عن خط الانحدار التقديري.

- ايجاد الخطأ المعياري لميل انحدار النموذج، s_b

$$s_b = \frac{s_e}{\sqrt{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}}$$

$$= \frac{213.974}{\sqrt{63299.08 - \frac{(629.8)^2}{13}}} = \frac{213.974}{181.074} = 1.182$$

- اختبار فرضية من ان ميل انحدار المجتمع $\beta \le 0.5$ وعند مستوى معنوية $\alpha = 0.05$

- الفرضية

$$H_0 : \beta \le 0.5$$
$$H_1 : \beta \rhd 0.5$$

- من الملحق رقم (4.3) وعند مستوى معنوية $\alpha = 0.05$ ودرجات حرية $\upsilon = 11$ ولاختبار من جانب واحد فان $t = 1.796$ (الجدولية) - لدينا $b=0.291$, $sb = 1.182$ وايجاد قيمة t المحتسبة وهي :

$$t=\frac{b-\beta}{s_b}=\frac{0.291-0.5}{1.182}=0.176$$

وحيث ان قيمة t المحتسبة هي أقل من القيمة الجدولية، فيكون القرار هو قبول فرضية العدم H_0 والاستدلال على ان قيمة ميل انحدار المجتمع هي اقل من 0.5

◄ ايجاد مدى الثقة لميل انحدار المجتمع β عند ثقة مقدارها 90٪

$$b-t_{(1-\alpha/2)},v.s_b \le \beta \le b+t_{(1-\alpha/2)},v.s_b$$

$$0.291-(1.796)(1.182)\le\beta\le 0.291+(1.796)(1.182)$$

$$1.831\le\beta\le 2.413$$

◄ ايجاد الخطأ المعياري لقيمة متوسط المتغير التابع $s_{\bar{y}}$ عند $x = 75.5$

- يكون لدينا :

$$\hat{y}=189.539+(0.29)(75.5)=211.51$$

$$\bar{x}=\frac{\sum x_i}{n}=\frac{629.8}{13}=48.446$$

- فنحصل على :

$$s_{\bar{y}}=s_e\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum x^2-\frac{(\sum x)^2}{n}}}$$

$$=213.974\sqrt{\frac{1}{13}+\frac{(75.5-48.446)^2}{63299.08-\frac{(629.8)^2}{13}}}$$

$$=213.974\sqrt{\frac{1}{13}+\frac{731.919}{632909.08-30511.387}}=67.401$$

◄ ايجاد مدى الثقة لمتوسط المجتمع $\bar{Y}$ والمعامل الثابت α عند ثقة مقدارها 90٪

- بالاعتماد على نتائج المثال وعند $x = 75.5$ بالنسبة لمدى الثقة لـ $\bar{Y}$ نحصل على :

$$\bar{y} - t_{(1-\alpha/2),} v.s_{\bar{y}} \le \bar{Y} \le \bar{y} + t_{(1-\alpha/2),} v.s_{\bar{y}}$$

$$211.51 - (1.796)(67.401) \le \bar{Y} \le 211.51 + (1.796)(67.401)$$

$$90.465 \le \bar{Y} \le 332.555$$

- اما بالنسبة للمعامل الثابت α فنحصل على :

$$a - t_{(1-\alpha/2),} v.s_{\bar{y}} \le \alpha \le a + t_{(1-\alpha/2),} v.s_{\bar{y}}$$

$$189.539 - (1.796)(67.401) \le \alpha \le 189.539 + (1.796)(67.401)$$

$$68..494 \le \alpha \le 310.584$$

مع الاشارة الى ان فترة الثقة تصبح مساوية لكل من $\bar{Y}$ و α عندما $x = 0$

6-2-5 حالة دراسية $C_{6\text{-}1}$

استخدام برنامج SPSS في الانحدار الخطي البسيط

لمعرفة مدى تاثير متغير اجور النقل Tcsat على متغيرمستوى الرضا العام عن خدمات النقل العام Gsat موضوع المثال (1.1)، سيتم اخضاع معطيات المتغيرين الى تحليل الانحدار الخطي البسيط، وكالاتي :

- استدعاء قائمة Analysis ومنها الامر الفرعي Regression ومن ثم التاشير على خيار Linear

- يظهر لنا مربع الحوار Linear Regression المبين في الشكل البياني رقم (2.6)، وفيه يتم استحدام السهم الجانبي لنقل المتغير التابع Gsat الى تحت Dependent والمتغير المستقل Tcsat .
- الكبس على ايقونة Statistics لتظهر لنا لوحة Linear Regression: Statistics المبينة في الشكل البياني رقم (3.6) ليتم التاشير ازاء المعايير الوصفية المتعلقة بقياس معنوية النموذج ومعاملات النموذج وكما هو موضح على الشكل البياني. بعد الانتهاء مع لوحة Linear Regression: Statistics يتم الكبس على ايقونة Continue للعودة الى مربع الحوار ،
- الكبس على ايقونة Options للحصول على لوحة Linear Regression: Options المبينة في الشكل البياني رقم (4.6) فيتم التاشير عندها في حالة الرغبة في تغييرما هو مثبت من معايير ادخال المتغير للتحليل او حذفه وكما مبين في الشكل البياني المذكور . والعودة مرة اخرى الى مربع الحوار ،
- الكبس على ايقونة Plots لتظهرلنا لوحة Linear Regression: Plots المبينة في الشكل البياني رقم (5.6)، ليتم التاشير على الاشكال البيانية المرغوب الحصل عليها والتي تعطي فكرة عن انتشار المعطيات ومدى تحقق فرضية الخطية Linearity وعن مدى تجانس انتشار الاخطاء Residuals وعن شكل التوزيع الطبيعي للمعطيات، وما الى ذلك. مع ملاحظة ان يتم الكبس على ايقونة Next الموجودة في وسط اللوحة، بعد الانتهاء من تحديد الشكل الاول، ليتم تحديد الشكل الثاني وهكذا، وبعد الانتهاء من لوحة Linear Regression: Plots والعودة الى مربع الحوار ،

- الكبس على ايقونة Ok لنحصل على المخرجات المبينة في الجداول رقم (2.6) والاشكال البيانية رقم (6.6) و (7.6) و (8.6) .

ومن المخرجات نستدل على ان جميع المعايير الاحصائية وهي، F , t, R , R^2 عالية المعنوية عند $\alpha= 0.000$ و المعامل الثابت عند $\alpha=0.05$، مع تحقق الخطية كما مبين على الشكل البياني رقم (7.6)، وان انتشار الاخطاء متجانس (باستثاء قيمة واحدة يمكن معالجتها كقيمة شاذة ويتم ذلك اما بالحذف او التعديل) كما يتضح من شكل الانتشار رقم (8.6)، مع الاطمئنان بدرجة كافية من شكل التوزيع الطبيعي للمعطيات كما يتضح من المدرج التكراري رقم (6.6). وبذلك فان شكل النموذج ياخذ الشكل التالي :

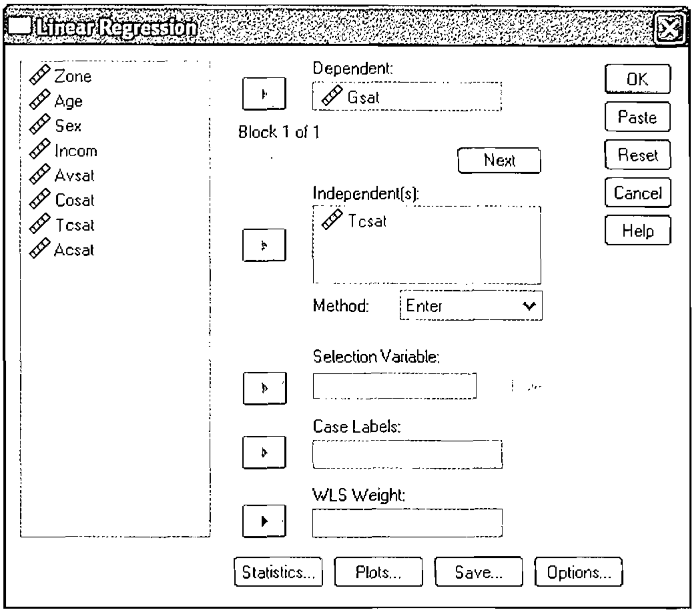
$$Gsat = 1.097 + 0.718\ Tcsat$$
$$R = 0.626 \quad , R^2 = 0.392$$
$$F = 18.67 \quad \text{Sig. at } 0.000$$

وبالنسبة للمغير المستقل لدينا :

$$T = 4.321 \quad \text{Sig. at } 0.000$$
$$\text{Std. Beta} = 0.626 \quad \text{Sig. at } 0.000$$

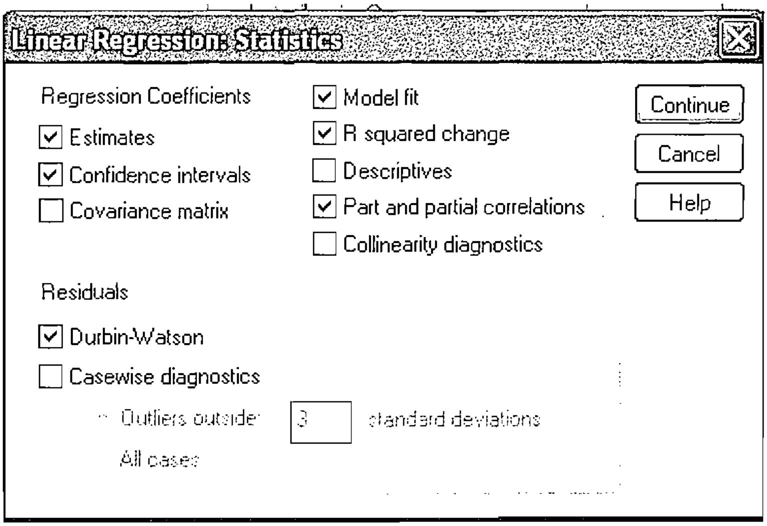
الشكل البياني رقم (2.6)

مربع حوار Linear Regression



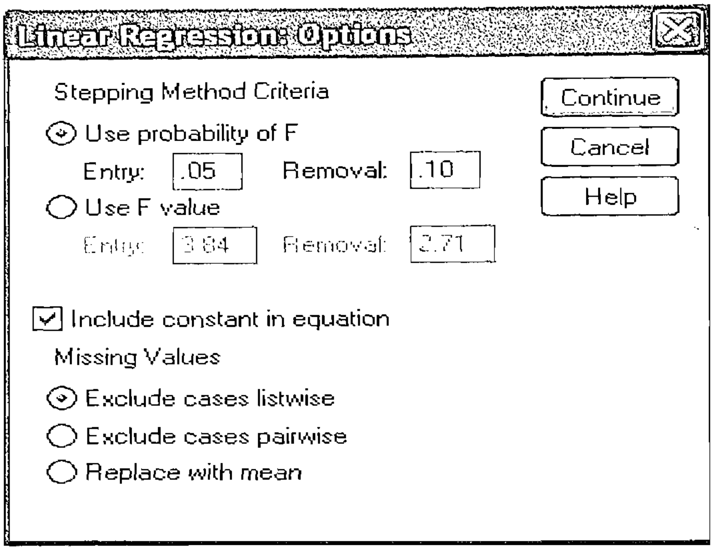
الشكل البياني رقم (3.6)

لوحة Linear Regression: Statistics



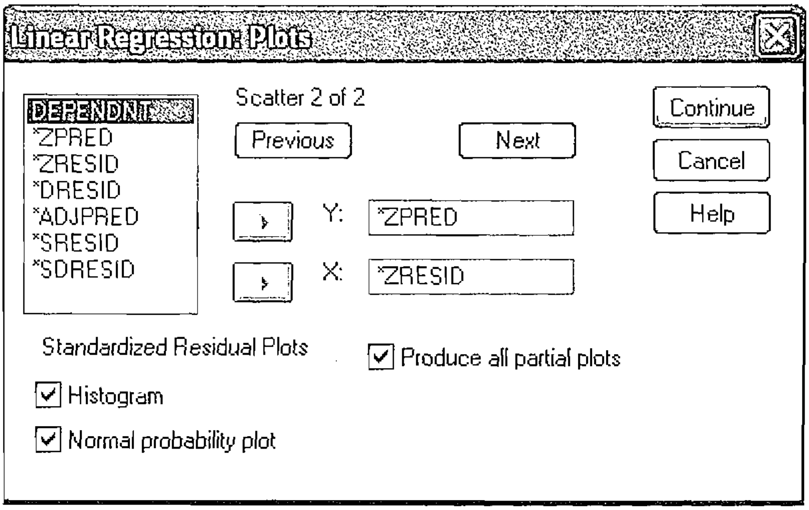
الشكل البياني رقم (4.6)

لوحة Linear Regression: Options



الشكل البياني رقم (5.6)

لوحة Linear Regression: Plots



جداول رقم (2.6)

مخرجات تحليل الانحدار الخطي البسيط

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	Tcsat(a)	.	Enter

a All requested variables entered.

b Dependent Variable: Gsat

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.626(a)	.392	.371	.703

a Predictors: (Constant), Tcsat

b Dependent Variable: Gsat

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9.223	1	9.223	18.672	.000(a)
	Residual	14.325	29	.494		
	Total	23.548	30			

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	Constant	1.097	.366		3.001	.005
	Tcsat	.718	.166	.626	4.321	.000

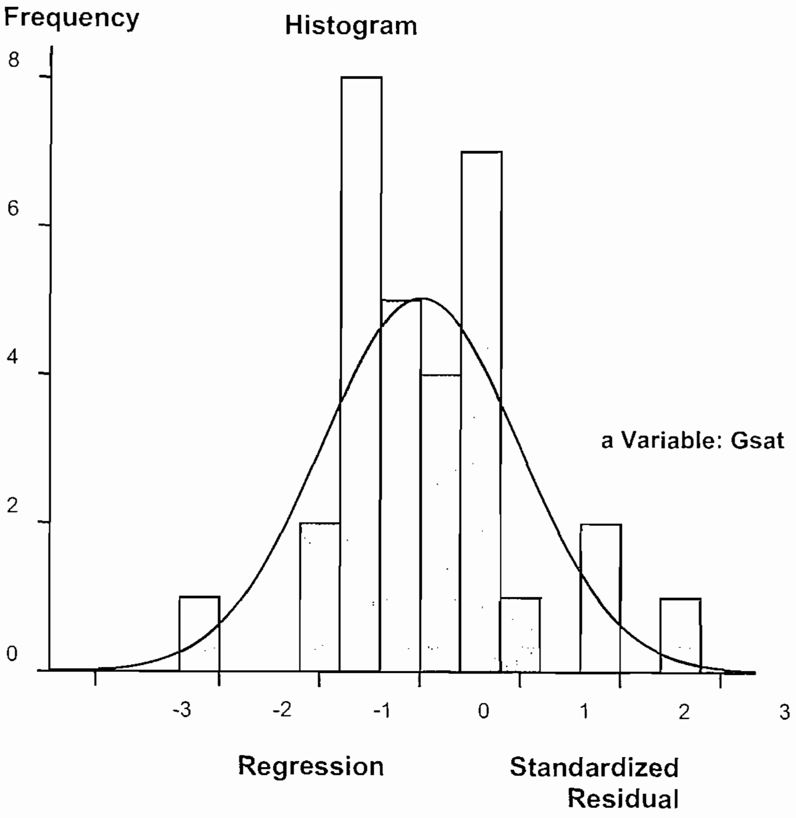
a Dependent Variable: Gsat

Residuals Statistics(a)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.82	3.97	2.58	.554	31
Residual	-1.534	1.747	.000	.691	31
Std. Predicted Value	-1.379	2.508	.000	1.000	31
Std. Residual	-2.183	2.486	.000	.983	31

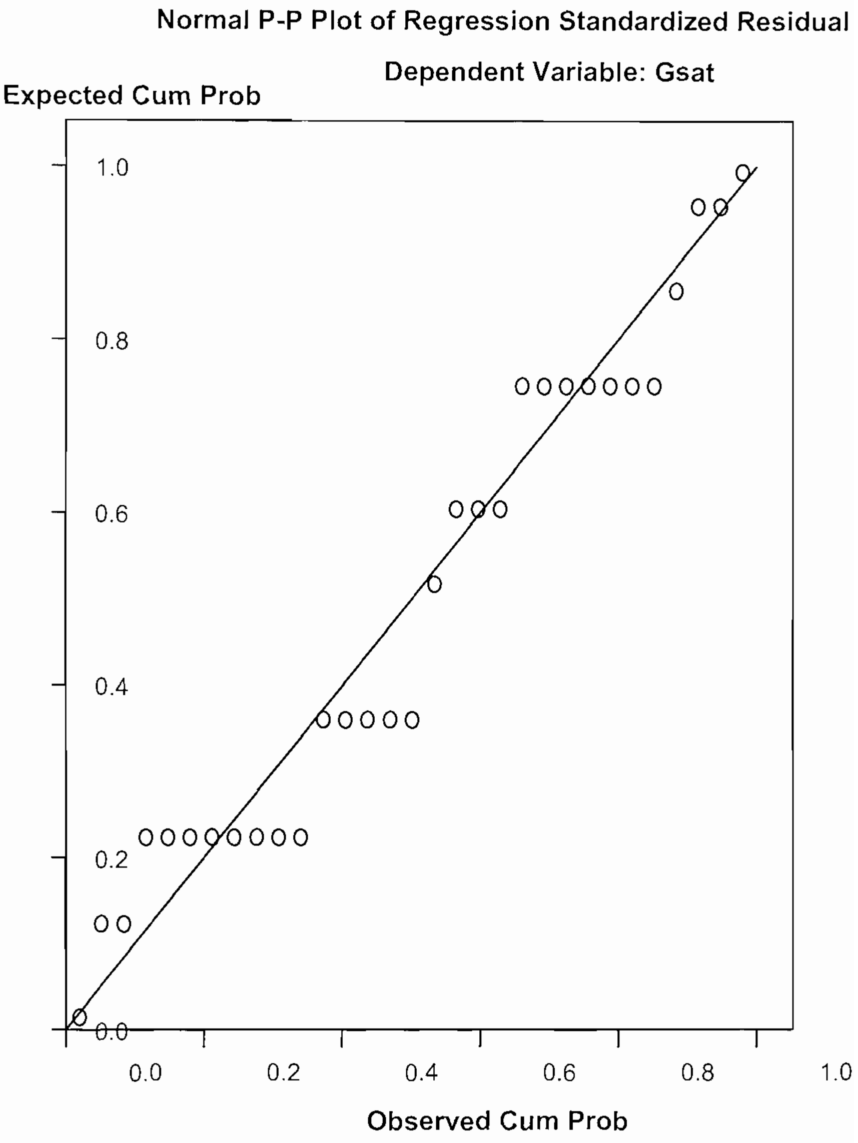
a Dependent Variable: Gsat

شكل بياني رقم (6.6)
التوزيع التكراري للقيم المعيارية للبواقي

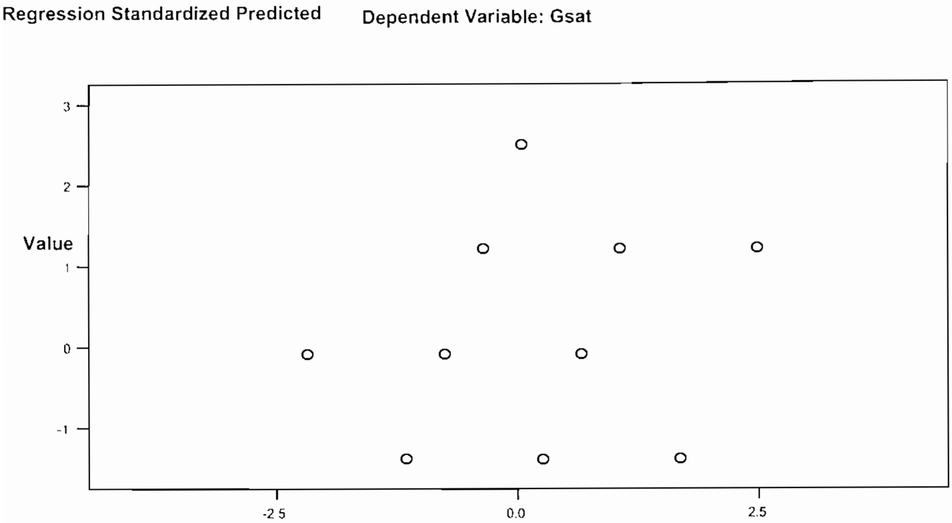


شكل بياني رقم (7.6)

يوضح العلاقة الخطية بين القيم المتوقعة والحقيقية



الشكل البياني رقم (8.6) شكل انتشار البواقي **Scatterplot**



3-6 الانحدار الخطي المتعدد Multiple Linear Regression

ويهدف استخدام تحليل الانحدارالمتعدد بصورة رئيسية البحث في العلاقة ما بين اكثر من متغير مستقل Independent Variables ويرمز لها X_i وتمثل العوامل المؤثرة على الظاهرة التي تكون تحت الدراسة، وبين المتغير التابع Dependent Variable ويرمز له Y والذي يمثل هذه الظاهرة سواء اكان البحث عن مدى تاثير مجموعة المتغيرات المستقلة او تاثير كل منها على حدة. ففي حالات عملية عديدة يكون المتغير التابع Y معتمدا في تفسيره على اكثر من متغير مستقل X فمثلا انتاج الحنطة (القمح) لايعتمد على المساحة الزروعة فقط بل ايضا على مستوى تسميد التربة و كمية المياه وعلى مكافحة الحشرات وغيرها، وان الطلب على القهوة لايعتمد على سعرها فقط بل على مستوى سعر الشاي ايضا وهكذا .

6-3-1 معادلة الانحدار الخطي المتعدد وطريقة تقدير β , α s'

ان المعادلة الخطية المتعددة هي الاساس لكافة الاشكال الاخرى للانحدار، والتي شكلها العام هو :

$$E(y) = \alpha + \beta 1 Xi1 + \beta 2 Xi2 + \ldots + \beta k X_i k + \varepsilon_i$$

حيث ان :

y_i = المتغير التابع (قيم المجتمع Y الحقيقية عند بناء النموذج)

X_i = المتغيرات المستقلة

ε_i = متغير الاخطاء العشوائية (البواقي)

α و β = المعامل الثابت ومعامل الانحدار على التوالي .

ويتم تقدير α و β باستخدام طريقة المربعات الصغرى حيث تصبح a وb هي رموز المعادلة التقديرية بدلا من α و β. فعند تضمين المعادلة لمتغيرين مستقلين، فأن معادلة الانحدار الخطية التقديرية لمعطيات عينة تاخذ شكل العلاقة التالية :

$$\hat{y} = a + b_1 x_{1i} + b_2 x_{2i} + e_i$$

حيث ان :

$$\sum e_i^2 = \sum (y_i - a - b_1 x_{1i} - b_2 x_{2i})^2$$
$$= \sum (y_i - \hat{y})^2$$

وتتم عملية التقدير لـ b_2, b_1, a وفقا لطريقة المربعات الصغرى وعلى غرار الخطوات التي تم اتباعها في حالة الانحدار الخطي البسيط، من خلال حل المعادلات المتعاقبة التالية :

$$\sum y_i^2 = na + b_1 \sum x_{1i} + b_2 \sum x_{2i}$$

$$\sum x_{1i}y_i = a\sum x_{1i} + b_1\sum x_{1i}^2 + b_2\sum x_{1i}x_{2i}$$
$$\sum x_{2i}y_i = a\sum x_{2i} + b_1\sum x_{1i}x_{2i} + b_2\sum x_{2i}^2$$

وان صيغ احتساب a , b_1 , b_2 المشتقة من طريقة المربعات الصغرى هي:

$$b_1 = \frac{\left[n\sum x_1 y - \sum x_1 \sum y\right]\left[n\sum x_2^2 - \left(\sum x_2\right)^2\right] - \left[n\sum x_2 y - \sum x_2 \sum y\right]\left[n\sum x_1 x_2 - \sum x_1 \sum x_2\right]}{\left[n\sum x_1^2 - \left(\sum x_1\right)^2\right]\left[n\sum x_2^2 - \left(\sum x_2\right)^2\right] - \left[n\sum x_1 x_2 - \sum x_1 \sum x_2\right]^2}$$

$$b_2 = \frac{\left[n\sum x_2 y - \sum x_2 \sum y\right]\left[n\sum x_1^2 - \left(\sum x_1\right)^2\right] - \left[n\sum x_1 y - \sum x_1 \sum y\right]\left[n\sum x_1 x_2 - \sum x_1 x_2\right]}{\left[n\sum x_1^2 - \left(\sum x_1\right)^2\right]\left[n\sum x_2^2 - \left(\sum x_2\right)^2\right] - \left[n\sum x_1 x_2 - \sum x_1 \sum x_2\right]^2}$$

$$a = \frac{\sum y - b_1\sum x_1 - b_2\sum x_2}{n}$$

ولاختبار معنوية معاملات الانحدار $b_1, b_2, \ldots, b_k$ يتم استخدام الاحصاءة t وكالاتي :

$$t = \frac{b_i}{s_{bi}}$$

حيث ان :

$$s_{b1} = \sqrt{\frac{\sum e_i^2}{n-k} \cdot \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 \sum x_2\right)^2}}$$

$$s_{b2} = \sqrt{\frac{\sum e_i^2}{n-k} \cdot \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 \sum x_2\right)^2}}$$

وبما ان :

$$R^2 = 1 - \frac{\sum e^2}{\sum y^2}$$

فان :

$$\sum e^2 = \sum y^2 (1 - R^2)$$

وبالحصول على قيم s_b, t يمكننا تحديد فترة الثقة لمعامل الانحدار الحقيقي للمجتمع β كالاتي :

$$b_i \pm t_i s_{bi}$$

6- 3- 2 معايير قياس كفاءة ومعنوية نموذج الانحدار الخطي المتعدد

(1) معايير احصائية Statistical Crireria

وتشمل t-test لاختبار معنوية معاملات المتغيرات المستقلة والعامل الثابت constant و r لاختبار درجة العلاقة بين كل متغير مستقل والمتغير التابع (dependent variable) ومنها ايضا F-ratio وR^2 لاختبار معنوية المعادلة النهائية ومدى معنوية درجة تفسير التباين. ويمكن اجمال اهم هذه المعايير الاحصائية بما يلي :

- **معامل التحديد (Coefficient of Determination)**

ويمثل النسبة المئوية للتباين التي يتم تفسيرها بواسطة المتغير او المتغيرات المستقلة التي يتضمنها النموذج. وهو يدل على مدى اقتراب المشاهدات من خط الانحدار. ويرمز لها بـ r^2 في حالة الانحدار الخطي البسيط وبـ في حالة الانحدار المتعدد R^2 وتقع قيمة R^2 بين 0 و 1 أي : $0 \le R^2 \le 1$ فكلما نقترب قيمة R^2 من 1 يعني ارتفاع معنوية النموذج التفسيرية. وصيغة حسابه كما في اعلاه هو :

$$R^2 = 1 - \frac{\sum e^2}{\sum y^2}$$

- **اختبار F (F- test)**

ويستخدم لاختبار معنوية المعادلة، بكلمة اخرى معنوية العلاقة بين مجموعة المتغيرات المستقلة والمتغير التابع Y، وكلما ارتفعت قيمة F الجدولية عند درجات حرية (n-k-1 , k) يكون قبولها بمعنوية اعلى، حيث ترمز n و k لعدد المشاهدات (العينة) وعدد المتغيرات المستقلة على التوالي. وصيغة اختبار F هي كما موضحة في اعلاه من هذا الفصل .

- **معامل الاتباط الجزئي r**

لاختبار درجة العلاقة بين كل متغير مستقل والمتغير التابع، وصيغة حسابه كما مبين في الفقرة (4.5) لموضوع تحليل الارتباط في الفصل الخامس.

- **اختبار t**

ويستخدم هذا المعيار لاختبار معنوية كل من معاملات الانحدار التي يتضمنها النموذج وذلك من خلال مقدار الخطا المعياري، S_{bi} وبواسطته يتسنى التعرف على مدى قابلية كل متغير مستقل على تفسير التذبذبات الحاصلة في المتغير التابع. وصيغته كما في اعلاه هي :

$$t = \frac{b_i}{S_{bi}}$$

(2) معايير منطقية Logical Criteria

وهي تخص الاشارة التي يجب ان يظهر معها معامل المتغير، ولكون القرار الذي يعتمد بشان صحة الاشارة او خطئها اساسه معرفة منطقية اتجاه

سلوك المتغير من حيث علاقته بالمتغير التابع لذا فقد سميت بالمعايير المنطقية، فعلى سبيل المثال بما ان انخفاض سعر الخدمة او السلعة يؤدي الى زيادة حجم الطلب، فمنطقيا يجب ان تظهر اشارة معامل المتغير سالبة، و حيث ان سهولة الوصول (Accessbility) للخدمة او السلعة يزيد من رضى الزبون، فمنطقيا ان تظهر الاشارة لمعامل متغير الوصول الى الخدمة او السلعة باشارة موجبة وهكذا

(3) الفرضيات Assumptions

وتتمثل بالتحقق من توزيع البواقي residuals كونها موزعة توزيعا طبيعيا واتجاهها خطيا للتاكد من عدم الحصول على تقديرات متحيزة وغير كفوءه، ويتم عادة التحقق من هذه الفرضيات من خلال الاشكال البيانية التي سيرد ذكرها. ويمكن اجمال اهم خصائص البواقي الازم التحقق منها بما يلي :

- ان وسطها الحسابي يساوي صفر، أي $E(\varepsilon_i) = 0$
- ان تباينها متساوي لكافة المشاهدات، أي $E(\varepsilon_i) = \sigma^2$
- ان قيمها مستقلة عن بعضها، أي $E(\varepsilon_i,\varepsilon_j) = 0$

حيث ان ε_i ترمز الى البواقي (residuales).

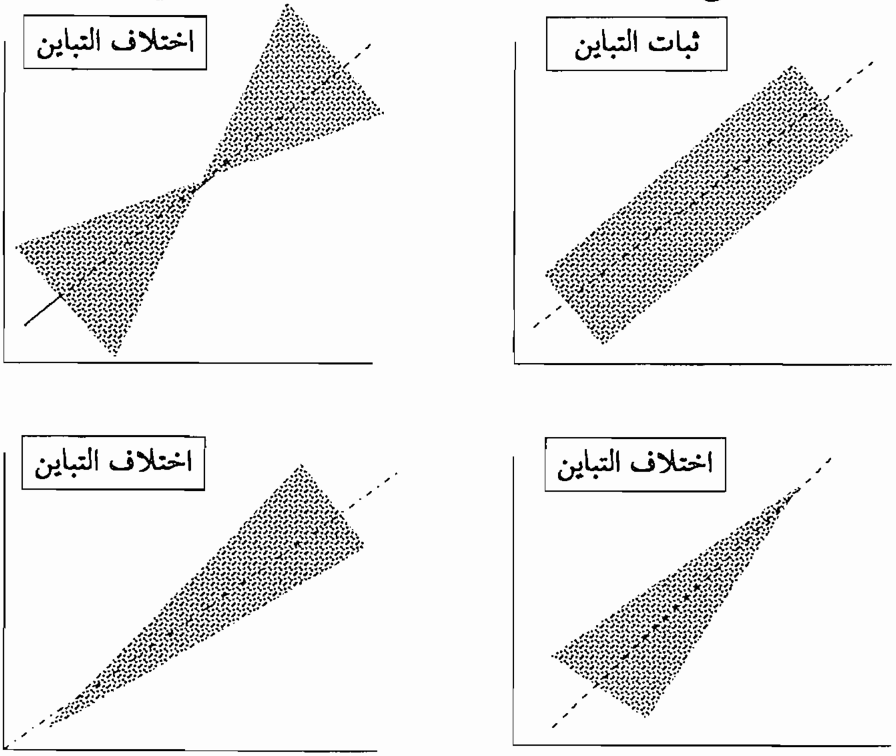
وهناك عدة طرق يمكن الاستعانة بها للتحقق من هذه الفرضيات والتي سيتم تناولها هنا، الا ان اهمها واكثرها استخداما هي طريقة توظيف الاشكال البيانية التالية (والملاحق من1.4الى 4.4 تعطي نماذج لهذه الاشكال البيانية) وهي :

- **الانتشار الخطي** للقيم حول خط دالة ميل الانحدار للتاكد من العلاقة الخطية بالنسبة للفرضية الاولى القائلة: ان الوسط الحسابي للبواقي يساوي صفر، أي $E(\varepsilon_i) = 0$

▪ **الانتشار المتجانس للفرضية الثانية:** ان تباين البواقي متساوية لكافة المشاهدات، أي $E(\varepsilon i) = \sigma^2$، وهي الفرضية التي يطلق عليها باختلاف التباين Heteroscedasticity ، وتظهرهذه المشكلة عند عدم ثبات تباين الخطأ العشوائي S_e^2 لقيم المتغيرات المستقلة، وبالتالي الحصول على قيم متحيزة وغير كفوءه، بكلمة اخرى اذا كانت قيمة الخطأ العشوائي تتغير بتغير قيمة e_i فتزداد بزيادة قيمة X_i، فمثلا اذا اخذنا عينة من الاسر حجمها n وكان التباين في الاستهلاك يزداد بارتفاع دخل الاسرة، فالاسرة التي دخلها اكبر يكون لديها مرونة اكبر في الاستهلاك. والاشكال التالية رقم (9.6) يمثل نماذج لحالات ثبات واختلاف تباين الخطأ العشوائي e_i.

شكل بياني رقم (9.6)

نماذج لحالات ثبات واختلاف تباين الخطأ العشوائي e_i



ومن طرق الكشف الاخرى عن اختلاف التباين بالاضافة لطريقة الاشكال البيانية هي: ترتيب قيم المعطيات من الاصغر فالاكبر للمتغير المستقل، واجراء نموذجين للانحدار، احدهما للقيم الصغيرة والاخر للقيم الكبيرة، وحذف لنقل 5 مشاهدات وسطية، فاذا كانت نسب مجموع مربعات الخطأ ESS للانحدار الثاني الى الانحدار الاول لمعرفة ان كانت قيمة F الجدولية تختلف معنويا مع درجات حرية (n-d-2k)/2، حيث ان : d ترمز لعدد المشاهدات المحذوفة و k عدد المعالم المقدرة. ولنتامل في النموذج التالي الذي يضم متوسط الاجور y وعدد العاملين X في 30 شركة لاحدى الصناعات، وان نتائج الانحدار هي :

$$\hat{y} = 7.5 + 0.009\,x$$
$$R^2 = 0.90$$

وبحذف المشاهدات لـ 6 الوسطى، وبناء نموذج للـ 12 مشاهدة الاولى واخر للـ 12 مشاهدة الثانية نحصل على :

$$\hat{y}_1 = 8.1 + 0.006\,x$$
$$R^2 = 0.66$$
$$ESS = 0.507$$

$$\hat{y}_2 = 6.1 + 0.01\,x$$
$$R^2 = 0.60$$
$$ESS = 3.095$$

وان :

$$\frac{ESS_2}{ESS_1} = \frac{3.909}{0.507} = 6.10$$

نجد ان قيمة $F_{(\nu=10,10)} = 2.97$ عند مستوى معنوية $\alpha = 0.05$ ومنها نستدل على وجود اختلاف في التباين .

وباعادة صياغة المتغيرات لتصحيح اختلاف التباين ،نحصل على النموذج التالي :

$$\frac{\hat{y}}{x} = 0.008 + 7.8\left(\frac{1}{x}\right)$$

$$R^2 = 0.99$$

نجد معامل ميل الانحدار اصبح 0.008 هو اصغر من 0.009 .

- **المدرج التكراري للتحقق من لفرضية الثالثة** لتوزيع البواقي لاثبات التوزيع الطبيعي للمعطيات واستقلالية المشاهدات التي تستخدم في الدراسة، أي $E(\varepsilon i,\varepsilon j) = 0$

كما ان فرضية استقلالية المشاهدات تظهر اليها الحاجة ايضا في حالة استخدام السلاسل الزمنية للتحقق من عدم وجود ارتباط ذاتي بين المشاهدات Autocorelation ويجري التحقق منها باستخدام صيغة Durbin - Watson والمبينة قيمها الجدولية في الملحق رقم (1.6) علما بان مشكلة الارتباط الذاتي لاتظهر في الدراسات التي تعتمد على بيانات مقطعية Cross sectional data. فعندما لاتتحقق هذه الفرضية، سيعني ذلك ظهور مشكلة الارتباط الذاتي Autocorrelation، اي ان المتغير العشوائي ε_i الذي يعود لفترة زمنية معينة يكون مرتبطا طرديا بالفترة الزمنية السابقة لها، وهو امر شائع في تحليل السلاسل الزمنية مما يؤدي الى التحيز نحو الاسفل، وبالتالي فان نتائج الاختبارات وفترات الثقة تكون مزيفة او خاطئة، وتستخدم طريقة Durbin- Watson عند مستوى معنوية معينة α ولعدد مشاهدات حجمها n و k ان كانت القيمة المحتسبة للمعادلة d المبينة في الاتي هي اصغر من القيمة الجدولية $d < d_L$ (d_L الحد الادنى) عندها نستدل على

وجود ارتباط ذاتي موجب، وبعكسها نرفض وجود الفرضية في حالة $d >$ d_U (d_U الحد الاعلى) .

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

وفي حالة الحاجة لتصحيح النموذج بسبب وجود الارتباط الذاتي، نقوم باستخدام المعامل y_{i-1} لتقدير معامل الارتباط الذاتي p وكالاتي :

$$\hat{y}_i = a(1-p) + py_{i-1} + b_1x_i + b_1x_{i-1} + e_i$$

ثم يعاد تقدير الانحدار على المتغيرات المحولة وكالاتي :

$$(y_i - \hat{p}y_{i-1}) = a(1-\hat{p}) + b_1(x_i - \hat{p}x_{i-1}) + (e_i - \overline{p}e_{i-1})$$

ولتجنب ضياع المشاهدة في عملية ايجاد الفروق نستخدم الصيغة التالية لكل من y و x للمشاهدة الاولى المحولة :

$y_1\sqrt{1-\hat{p}^2}$ و $x_1\sqrt{1-\hat{p}^2}$ على التوالي .

فمثلا لو كان النموذج التالي يمثل مستوى المخزون y والمبيعات x (بالمليار دولار) في احدى الصناعات التحويلية الامريكية، وان نتائج انحدار y على x هي :

$$\hat{y}_i = 6.61 + 1.61\, x_i$$
$$R^2 = 0.98 \quad , \quad d = 0.70$$

ومن الملحق رقم (1.6) نجد ان: $d_L = 1.20$ عند مستوى معنوية $\alpha = 0.05$ مع $n = 20$ و $k = 1$، وهي بذلك $d < d_L$ فنستدل على

وجود ارتباط ذاتي، ولتصحيح النموذج فان معامل P في الانحدار التالي يعطي تقديرا للمعامل y_{i-1}، اي :

$$\hat{y}_i = 4.08 + 0.74\ y_{i-1} + 1.49\ x_i - 1.11\ x_{i-1}$$

وباستخدام $\hat{p} = 0.74$ لتحويل المتغيرات الاصلية لكل من y و x للمشاهدة الاولى، باستخدام :

$$52.9\sqrt{1-(0.74)^2} = 35.58 \quad و \quad 30.3\sqrt{1-(0.74)^2} = 20.37$$

نحصل على :

$$\hat{y}_i = 4.14 + 1.49\ x_1$$
$$R^2 = 0.92 \quad , \quad d = 1.46$$

وباستخدام الملحق (1.6) نجد ان: $d = 1.46 > d_U = 1.41$ اي ليس هناك دليل على وجود الارتباط الذاتي .

- **بالاضافة لما سبق فهناك** حاجة ايضا للتحقق من فرضية عدم وجود علاقات متداخلة (Multicolinearity) بين المتغيرات المستقلة فيتم التحقق منها من خلال استخدام مصفوفة الارتباط وكما اشرنا لذلك في اعلاه او بتوظيف تحليل المركبات الاساسية Principal Componenet Analysis كما سنلاحظ في المواضيع اللاحقة .

وتحصل هذه الحالة عندما يكون اثنين اوكثر من المتغيرات المستقلة التي يضمها النموذج على ارتباط قوي، مما يجعل من الصعب تحديد تاثير كل من هذه المتغيرات على المتغير التابع، وبالتالي فان معاملات الانحدار b's غير معنوية أحصائيا وقد تاتي باشارات خاطئة ايضا رغم معنوية معامل الارتباط R ومعامل التحديد R^2.

وللتغلب على هذه المشكلة يتم التخلص من واحد او اكثر من المتغيرات ذات الارتباط العالي، او بزيادة حجم العينة او اللجوء الى تحويل صيغة المتغيرات كأن تصبح لوغارتيمية او نصف لوغارتيمية او غيرها .

والنموذج التالي الذي يرمز فيه الى الايرادات بـ y و x_1 و بـ x_2 للرقم القياسي لاسعار المستهلكين وجميعها بمليارات الدولارات للولايات المتحدة الامريكية (المؤلف-1997)، يوضح بانه رغم ان R^2 هي عالية المعنوية، الا ان قيم كل من b_1 و b_2 هي غير معنوية احصائيا عند 0.05، والسبب واضح عند الرجوع الى معامل الارتباط $r_{12} = 0.997$.

$$\hat{y} = -101.49 + 0.08\,x_1 + 0.76\,x_2$$

Sig. at : 1.400 1000

$$R^2 = 0.985 \quad , r^2 = 0.997$$

وفي حالة استبعاد x_2 نحصل على :

$$\hat{y} = -69.03 + 0.13\,x_1$$

Sig. at : 0.01

$$R^2 = 0.98$$

مع الاشارة الى ان قرار استبعاد اي من المتغيرين يعتمد على المعايير الاحصائية ودرجة اهمية المتغير بالنسبة للظاهرة وعلى كلفة القياس .

(4) اختبار القوة التنبوئية للنموذج

Predictive Power of the Model

وفي هذا الاختبار يتم تقييم مدى قدرة طاقم المتغيرات التي يتضمنها النموذج على تقدير قيم لا تختلف جوهريا عن القيم الحقيقية للمتغير التابع. وتتم عملية التقيم من خلال اختبار الفروق الناتجة بين القيم الحقيقية (y)

والقيم التي يتم تقديرها بواسطة النموذج ($\hat{y}$)، ومن ان حجم الفروق المعيارية لاتتجاوز مقدار الخطا المسموح. وهناك عدة طرق يمكن توظيفها لهذا الغرض وجميعها تفترض بان هذه الفروق موزعة توزيعا طبيعيا، ومنها طريقة الانحرافات الطبيعية (Normal Deviates)، وطريقة البواقي المعيارية (Standarized Residuals) وجميعها تفنرض وقوع هذه البواقي المعيارية بين حدي 1.96- و1.96+ عند درجة ثقة مقدارها 95٪ وان الشكل العام لصيغة طريقة الانحرافات الطبيعية هي :

$$ND = e_i / s$$

حيث ان :

$$e_i = y - \hat{y}$$

$$s = \frac{\sqrt{\sum e_i^2}}{n - k - 1}$$

ويتم بيانيا وكما هو في الشكل رقم (10.6)، توضح مدى تقارب القيم الحقيقية للمتغير التابع مع القيم التي يتم استخراجها بواسطة النموذج الذي يتم تطويره من خلال حجم الفروق (البواقي القياسية للانحدار) عند درجة ثقة 95٪ .

(5) الاختبار العملي للنموذج المطور

Practical Testing of Developed Model

وللزيادة في التاكد من جودة وفعالية النموذج بعد ان يتم التحقق من استيفاءه لكافة المعايير والفرضيات التي اشرنا اليها في اعلاه، يمكن القيام بتقسيم عينة المعطيات التي استخدمت في بناء النموذج الى قسمين وتطبيق النموذج المطور(الذي تم بناءه) على كل قسم منها لمعرفة مدى تقارب قيم المعاملات الناتجة مع النموذج الاصلي وكذا مع معايير الجودة والتحقق من فرضيات كل منها وقبول نتائج الاختبار عند ثقة مقدارها 95 ٪ .

6- 3- 3 طرق الانحدار الخطي المتعدد

هناك عدة طرق للانحدار التي يتم توظيفها لاختيار افضل طاقم للمتغيرات المستقلة لتضمينها في النموذج الذي يتم بناؤه ان جوهر الافكار التي تعتمد عليها جميع طرق الاختيار التي سيلي ذكرها هي تضمين المتغير الذي يضيف اكبر زيادة ممكنة الى قوة التفسير للنموذج، واذا كان على المتغير ان يحذف فيجب ان يكون تأثير حذفه اقل ما يمكن على قدرة النموذج التفسيرية. أما أهم طرق الانحدار المتعدد المستخدمة لاختيار افضل طاقم متغيرات مستقلة فهي:

(1) طريقة شمول كافة المتغيرات All Possible Regression

وتستخدم اذا كان عدد المتغيرات ليس كبيرا، وابرز عيوبها حاجتها لعمليات حسابية ووقت كبيرين.

(2) طريقة الاضافة المتتالية Forward Selection Regression

وفيها اذا كانت قيمة F المجدولة هي اقل من المحتسبة عندها يتوقف البحث عن متغير، وبعكسه يتم ادخال متغير جديد الى المعادلة واعادة الاحتساب , أي :

$$H_0: \beta_i = 0$$
$$\text{vs.} \quad H_1: \beta_i \neq 0$$

(3) طريقة الحذف التنازلي

Backward Elimination Selection Regression

وهذا اذا كانت قيمة F المحتسبة لكافة المتغيرات اكبر من قيمة F الجدولية، عندها يحذف متغير من المعادلة والرجوع لمعرفة قيمة F المحتسبة من جديد وهكذا لغاية تفوق قيمة F الجدولية.

(4) طريقة الخطوات المتتالية

Stepwise Selection Regression

تجمع بين طريقتي الاضافة المتتالية (FS) والحذف التنازلي (BE)، وفي كل خطوة يتم اختيار متغير ابتداء من الاكثر اهمية ولغاية عدم هبوط قيمة F المحتسبة عن قيمة F الجدولية بكلمة اخرى اجراء اختبار معاملات المتغيرات لمعرفة معنويتها من عدمها.

وتعتبر طريقة الخطوات Stepwise Regression هي اكثر الطرق استخداما وانتشارا من الناحية العملية لقلة الوقت الذي تحتاجه في عملية الاحتساب بالإضافة الى انها تعرض النتائج في كل خطوة بصورة واضحة ومرضية ومبكرة من دون الحاجة لاجراء الخطوات غير المعنوية.

6-3-4 حالة دراسية C_{6-2}

استخدام برنامج SPSS في تحليل الانحدار الخطي المتعدد

سيتم هنا توظيف ذات المعطيات التي تم استخدامها في الحالة الدراسية المتعلقة بتوظيف الانحدار في التحليل الوصفي موضوع الفقرة (5-3) والفقرة (5-4) ايضا، والمتعلقة بالعوامل المؤثرة على الانتاج البحثي والتي تشمل عينة عددها 74 من التدريسيين في الجامعات، والمتغيرات التي سيتم اخضاعها للتحليل يبلغ عددها 26 متغيرا، 10 منها تخص خصائص المبحوثين و 15 متغيرا تتعلق بمستوى رضاهم عن الظروف المحيطة بمجال البحوث والدراسات في الجامعات التي يعملون بها، بالاضافة للمتغيرالتابع وهو عدد المؤلفات والبحوث المنشورة تعبيرا عن الانتاج البحثي. وذلك بهدف بناء نموذج احصائي يضم العوامل ذات التاثير على الانتاج البحثي في جامعات عربية، لاجل توفير أداة علمية تساعد في تخطيط لتطوير عملية البحث العلمي

(1) ان اجراءات تحليل الانحدار الخطي المتعدد باستخدام برنامج SPSS هي ذات الاجراءات التي تم اتباعها مع حالة التحليل الخطي البسيط في اعلاه. مع التاكيد هنا على نقطيتين هما :

الاولى : هي ضرورة اختيار طريقة التحليل Method الموجودة في مربع الحوار المبين شكله في (2.6) اعلاه، ويفضل اختيار طريقة Stepwise عندما يكون الهدف الحصول على نموذج لاغراض التنبوء او التقدير او السيطرة والتحكم، لان الطريقة وكما سبق الاشارة في اعلاه، تساعد على الاختصار في الوقت من جهة وتقوم بعرض نتائج كل متغير يتم اضافتة لعملية التحليل من جهة اخرى. اما اذا كان الهدف من التحليل هو الوصف والتفسير للظاهرة تحت الدراسة، عندها يفضل التاشير على طريقة Enter لانها ستقوم بشمول كافة المتغيرات في عملية التحليل وبناء النموذج الوصفي وكما اشرنا لذلك في فقرة التحليل الوصفي.

اما النقطة الثانية، هو التاكيد على ضرورة التاشير على الرسوم البيانية في لوحة Plots والمبين شكلها في (5.6) اعلاه ايضا، لغرض التحقق من الفروض وعلى الاخص تلك المتعلقة بالخطية والتوزيع الطبيعي وشكل انتشار الاخطاء.

(2) عقب اخضاع ملف المعطيات الذي تم انشاؤه، واتباع ذات الاجراءات المبينة في تحليل الانحدار الخطي البسيط في الفقرة (6-2-5) اعلاه، جاءت مخرجات نتائج التحليل والمبينه في جداول المخرجات رقم (3.6) والاشكال البيانية المبينة رقم (10.6) و (11.6) و (12.6) . ومنها نستل على ظهور5 متغيرات من مجموع 25 متغيرا مستقلا، مستوفية لمعايير المعنوية وكما مبين من جداول المخرجات رقم (3.6)، وهي :

Tit (اللقب العلمي)

Nay (فئات سنين الخدمة الاكاديمية)

Age (العمر)

Spe (الاختصاص العلمي)

X_{14} (عدم جدية الجهات المستفيدة في تطبيق نتائج البحوث)

وجاءت معاملات الانحدار لكل من هذه المتغيرات والمعامل الثابت ودرجة معنويتها، ومعايير جودة النموذج من خلال كل من F, R^2 & R والمبينة في مجموعة الجداول التالية رقم (3.6) كما يلي :

Coeffic.	Variable	S.E	T	Sig.
ŷ=8.634	(Constnat)	1.863	4.634	0.000
-2.234	Tit	0.399	-5.602	0.000
+0.999	Nay	0.452	2.209	0.031
+1.324	Age	0.462	2.869	0.006
-0.357	Spe	0.135	-2.650	0.010
+0.458	X14	0.185	2.472	0.016

$R = 0.928$, $R^2=0.861$

F= 81.341 Sig. At 0.000

(3) مناقشة المعايير المطلوب تحققها للوقوف على مدى استيفاء النموذج لها بما في ذلك فحص الاشكال البيانية المتعلقة بمسالة الفرضيات التي تعتبر حاسمة عند استخدام النموذج في بناء التوقعات والتقديرات والخروج من هذه المناقشة عن قرار مدى صلاحية النموذج للاستخدام في العملية التخطيطية واتخاذ القرارات، وكالاتي :

▪ المعايير المنطقية Logical Criteria

بالرجوع الى الاشارات التي جاءت بها كل من المتغيرات التي ضمها النموذج الذي تم تطويره نجد بان جميعها جاءت صحيحة، فاشارة المتغير Tit سالبة هي نتيجة اعطاء القيمة الاقل وهي 1 للقب استاذ و 2 للاستاذ المشارك وهكذا، وبالتالي فمن المتوقع بانه كلما انخفضت قيمة المتغير يزداد الانتاج البحثي اي ترتفع قيمة المتغير التابعy.وكذا الحال عن الاشارة السالبة للمتغير Spe الذي يبدا بالاختصاصات العلمية التي اعطيت لها القيمة 1 وتاخذ بالتصاعد لغاية القيمة 6 مما يعني بان هذه الاختصاصات هي الاكثر انتاجا نتيجة اليسر في توفير المختبرات والاجهزة المطلوبة وما الى ذلك. اما اشارات المتغيرات الاخرى التي ضمها النموذج وهي Nay, Age,X_{14} فهي موجبة وجاءت متماشية ايضا مع صيغة طرحهما على المبحوث، فكل زيادة في معدل عدد سنين الخدمة الاكاديمية تؤول الى زيادة في الانتاج البحثي , وكذا الحال بالنسبة لمتغير العمر، اما بالنسبة لمتغير مستوى الرضا عن تطبيق نتائج البحوث من قبل الجهات المستفيدة، فزيادة مستوى الرضا التي يبدأ من 1 وترتفع لغاية 5 من شانها ان تؤدي الى ارتفاع في قيمة y المعبر عن عدد المؤلفات والبحوث المنشورة.

▪ المعايير الاحصائية Statistical Criteria

وبملاحظة مستوى معنوية المعاييرالتي ظهر بها النموذج سواء مايتعلق بمعامل التحديد (Coefficient of Determination, R^2) او أختبار F (F- test) او اختبار t نجد ان جميعها عالية معنوية highly significant واغلبها جاء عند مستوى 0.000 او 0.00 .

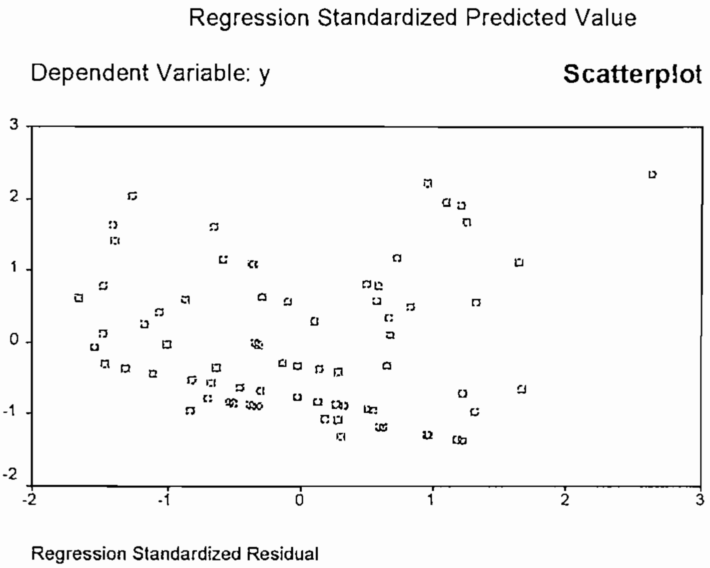
▪ اختبار فرضيات النموذج Assumptions

ان الاشكال البيانية لكل من الارقام (10.6) المتعلقه بفرضية العلاقة الخطية والذي يخص اختبار فرضية مساواة الوسط الحسابي للصفر $E(\varepsilon_i)=0$ و (11.6) المتعلق بشكل انتشار البواقي للتحقق من فرضية $E(\varepsilon_i)= \sigma^2$، والتوزيع الطبيعي للمدرج التكراري في الشكل (12.6) الذي يتعلق باختبار استقلالية البواقي وتوزيعها الطبيعي $E(\varepsilon_i,\varepsilon_j) = 0$. تعطي صورة واضحة عن استيفاء النموذج لكل من الفرضيات الثلاث.

شكل بياني رقم (10.6)

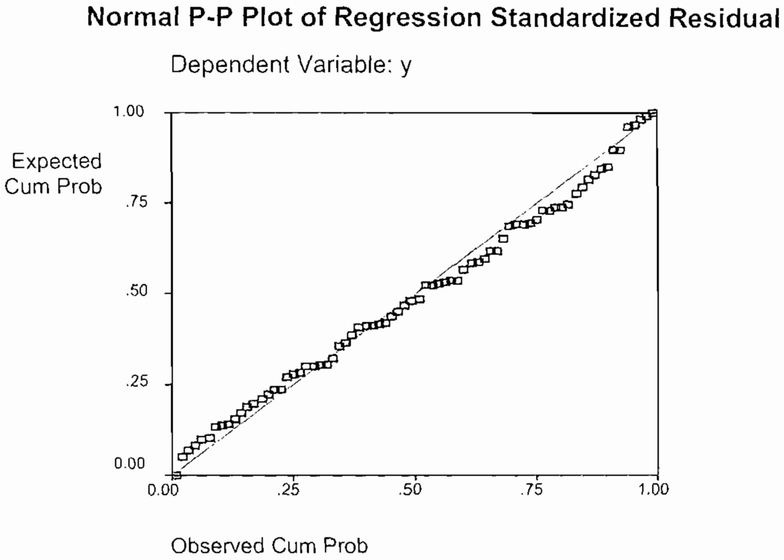
اختبار فرضية الاتجاه الخطي

ومساواة الوسط الحسابي للصفر، أي $E(\varepsilon_i) = 0$



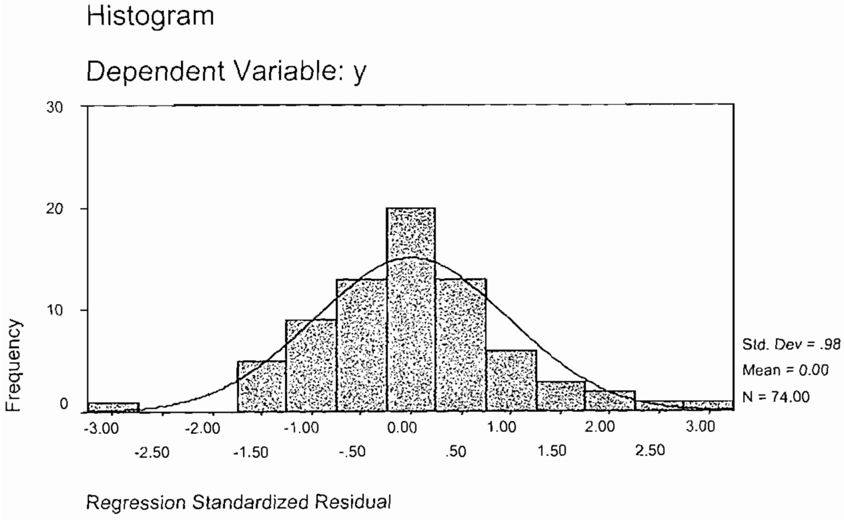
شكل بياني رقم (11.6)

يبين اختبار فرضية تساوى التباين لكافة المشاهدات، أي $E(\varepsilon_i) = \sigma^2$



شكل بياني رقم (12.6)

يبين نتيجة اختبار فرضية ان قيم البواقي مستقلة عن بعضها ، أي $E(\varepsilon i,\varepsilon j) = 0$



▫ اختبار القوة التنبوئية للنموذج

Predictive Power of the Model

وكما ذكرنا انفا يتم تقييم مدى قدرة طاقم المتغيرات التي يتضمنها النموذج على تقدير قيم لا تختلف جوهريا عن القيم الحقيقية للمتغير التابع. وتتم عملية التقيم من خلال اختبار الفروق الناتجة بين القيم الحقيقية (y) والقيم التي يتم تقديرها بواسطة النموذج ($\hat{y}$)، ومن ان حجم الفروق المعيارية لاتتجاوز مقدار الخطا المسموح. وهناك عدة طرق يمكن توظيفها لهذا الغرض وجميعها تفترض بان هذه الفروق موزعة توزيعا طبيعيا، ومنها طريقة الانحرافات الطبيعية(Normal Deviates)، وطريقة البواقي المعيارية (Standarized Residuals) وجميعها تفترض وقوع هذه البواقي المعيارية بين حدي 1.96- و1.96+ عند درجة ثقة مقدارها 95٪ .

والجدول رقم (4.6) التالي يعطي صورة عن تحليل البواقي القياسية لقيم التنبوء بواسطة نموذج الانحدار الذي تم تطويره لدرجة ثقة 95٪ ومقدارقيمتها الجدولية عند $\alpha/2 = 2.576$. مقابل القيمتين الدنيا والعليا -1.402 و 2.403 على التوالي والتي تقل عن القيمة الجدولية 2.576.

جدول رقم (4.6) مؤشرات تحليل البواقي

Residuals Statistics

	Minimum	Maximum	Mean	Std.Deviation	N
Predicted Value	.3409	17.1686	6.7753	4.5964	74
Residual	-4.5972	4.9012	5.91E-02	1.8218	74
Std. Predicted Va	-1.402	2.403	.053	1.039	74
Std. Residual	-2.491	2.655	-.032	.987	74

جداول رقم (3.6)

مخرجات تحليل الانحدار الخطي المتعدد باستخدام برنامج SPSS

Model Summary(j)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.873(a)	.763	.760	2.385
2	.923(b)	.852	.848	1.897
3	.933(c)	.870	.864	1.792
4	.940(d)	.883	.876	1.710
5	.944(e)	.892	.884	1.658
6	.943(f)	.890	.883	1.661
7	.948(g)	.898	.891	1.606
8	.952(h)	.906	.897	1.560
9	.955(i)	.911	.902	1.524

a Predictors: (Constant), ays
b Predictors: (Constant), ays, titl
c Predictors: (Constant), ays, titl, tys
d Predictors: (Constant), ays, titl, tys, Deg
e Predictors: (Constant), ays, titl, tys, Deg, Cou
f Predictors: (Constant), ays, tys, Deg, Cou
g Predictors: (Constant), ays, tys, Deg, Cou, x05
h Predictors: (Constant), ays, tys, Deg, Cou, x05, x14
i Predictors: (Constant), ays, tys, Deg, Cou, x05, x14, Spe
j Dependent Variable: y

ANOVA(j)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1317.514	1	1317.5	231.6	.000(a)
	Residual	409.527	72	5.68		
	Total	1727.041	73			
2	Regression	1471.548	2	735.77	204.6	.000(b)
	Residual	255.492	71	3.6		
	Total	1727.041	73			
3	Regression	1502.275	3	500.7	155.9	.000(c)
	Residual	224.766	70	3.2		
	Total	1727.041	73			
4	Regression	1525.230	4	381.3	130.4	.000(d)
	Residual	201.811	69	2.9		

	Total	1727.041	73			
5	Regression	1540.005	5	308.0	111.9	.000(e)
	Residual	187.036	68	2.75		
	Total	1727.041	73			
6	Regression	1536.677	4	384.2	139.3	.000(f)
	Residual	190.364	69	2.76		
	Total	1727.041	73			
7	Regression	1551.603	5	310.3	120.4	.000(g)
	Residual	175.438	68	2.58		
	Total	1727.041	73			
8	Regression	1564.066	6	260.68	107.2	.000(h)
	Residual	162.975	67	2.43		
	Total	1727.041	73			
9	Regression	1573.767	7	224.8	96.8	.000(i)
	Residual	153.273	66	2.32		
	Total	1727.041	73			

a Predictors: (Constant), ays
b Predictors: (Constant), ays, titl
c Predictors: (Constant), ays, titl, tys
d Predictors: (Constant), ays, titl, tys, Deg
e Predictors: (Constant), ays, titl, tys, Deg, Cou
f Predictors: (Constant), ays, tys, Deg, Cou
g Predictors: (Constant), ays, tys, Deg, Cou, x05
h Predictors: (Constant), ays, tys, Deg, Cou, x05, x14
i Predictors: (Constant), ays, tys, Deg, Cou, x05, x14, Spe
j Dependent Variable: y

Coefficients(a)

Model		Unstandardized Coefficients		Sd. Coeff	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Low Bound	Upp Bound
1	Constant	1.210	.456		2.655	.010	.302	2.119
	ays	.567	.037	.873	15.220	.000	.493	.642
2	Const	10.714	1.497		7.156	.000	7.729	13.699
	ays	.319	.048	.490	6.602	.000	.222	.415
	titl	-2.510	.384	-.486	-6.543	.000	-3.275	-1.745
3	Constant	9.443	1.473		6.412	.000	6.506	12.381
	ays	.191	.061	.294	3.103	.003	.068	.313
	titl	-2.206	.375	-.427	-5.874	.000	-2.954	-1.457
	tys	.132	.043	.280	3.093	.003	.047	.217

4	Constant	8.495	1.446		5.876	.000	5.611	11.380
	ays	.228	.060	.351	3.789	.000	.108	.348
	titl	-1.071	.541	-.207	-1.981	.052	-2.150	.008
	tys	.151	.041	.319	3.650	.001	.068	.233
	Deg	-1.866	.666	-.193	-2.802	.007	-3.195	-.537
5	Constant	8.975	1.417		6.333	.000	6.147	11.803
	ays	.234	.058	.361	4.012	.000	.118	.351
	titl	-.616	.560	-.119	-1.100	.275	-1.733	.501
	tys	.159	.040	.336	3.948	.000	.078	.239
	Deg	-1.934	.647	-.200	-2.990	.004	-3.224	-.643
	Cou	-.718	.310	-.112	-2.318	.023	-1.336	-.100
6	Constant	7.898	1.025		7.702	.000	5.852	9.943
	ays	.260	.053	.401	4.871	.000	.154	.367
	tys	.172	.038	.364	4.483	.000	.095	.248
	Deg	-2.443	.452	-.253	-5.408	.000	-3.345	-1.542
	Cou	-.837	.290	-.131	-2.883	.005	-1.417	-.258
7	(Constant	10.086	1.346		7.494	.000	7.401	12.772
	ays	.257	.052	.396	4.979	.000	.154	.361
	tys	.176	.037	.373	4.746	.000	.102	.250
	Deg	-2.663	.446	-.275	-5.966	.000	-3.554	-1.772
	Cou	-.858	.281	-.134	-3.052	.003	-1.418	-.297
	x05	-.641	.266	-.097	-2.405	.019	-1.172	-.109
8	(Constant	9.119	1.375		6.632	.000	6.374	11.863
	ays	.242	.051	.372	4.775	.000	.141	.343
	tys	.187	.036	.395	5.133	.000	.114	.259
	Deg	-2.528	.438	-.261	-5.777	.000	-3.401	-1.654
	Cou	-.832	.273	-.130	-3.047	.003	-1.377	-.287
	x05	-.648	.259	-.098	-2.506	.015	-1.164	-.132
	x14	.353	.156	.088	2.264	.027	.042	.664
9	(Constant	10.085	1.424		7.081	.000	7.242	12.929
	ays	.233	.050	.358	4.678	.000	.133	.332
	tys	.188	.036	.399	5.306	.000	.118	.259
	Deg	-2.445	.429	-.253	-5.693	.000	-3.302	-1.587
	Cou	-.862	.267	-.135	-3.225	.002	-1.395	-.328
	x05	-.712	.255	-.107	-2.797	.007	-1.220	-.204
	x14	.367	.153	.091	2.404	.019	.062	.671
	Spe	-.223	.109	-.077	-2.044	.045	-.440	-.005

a Dependent Variable: y

Residuals Statistics(a)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.01	18.01	6.72	4.643	74
Residual	-3.300	3.699	.000	1.449	74
Std. Predicted Value	-1.444	2.432	.000	1.000	74
Std. Residual	-2.165	2.427	.000	.951	74

a Dependent Variable: y

4-6 تحليل حساسية النموذج Sensitivity Analysis

6- 4- 1 مفهوم وخصائص تحليل الحساسية

وهي عبارة عن اجراءات يمكن بواسطتها تصور الحال الذي سيؤول اليه المتغير التابع (الانتاج البحثي) من خلال افتراضات تتعلق بكل من المتغيرات المستقلة التي يتضمنها النموذج باعتماد درجة مرونة كل من هذه المتغيرات المستقلة والحدود التي يمكن ان يذهب اليها في عملية التطوير. ان هذا النوع من التحليل يتيح لمتخذ القرار او المخطط وضع عدة سيناريوهات ويوفر بدائل ليختار الافضل من بينها وبما يتناسب و امكانياته المتاحة. بكلمة اخرى ان عملية تحليل الحساسية ترينا درجة استجابة المتغير التابع للتغير الذي يمكن ان يطرأ على أي من المتغيرات المستقله للنموذج. وان الصيغة التي يمكن استخدامها لقياس درجة مرونة المتغير المستقل الذي يتم توظيفه في تطوير المتغير (y) هي :

$$E_i = b_i \quad \frac{\overline{X}_i}{\overline{y}}$$

حيث يكون الاستدلال على ان المتغير عالي المرونة عندما تكون النتيجة هي $E_i > 1$ ومرن عند $E_i = 1$ وغير مرن عندما $E_i < 1$.

مع الاشارة هنا الى ضرورة ان تكون **وحدة قياس** كل من البسط والمقام موحدة، اي توحيد صياغة المتغيرات المستقلة والمتغيرالتابع عند اخضاعها لتحليل المرونات لتاتي متوسطاتها $\bar{x}$ و $\bar{y}$ متوائمة القياس للحصول على نتائج صحيحة في تحليل المرونات. ولاجل تماشى المتغيرات المستقلة التي تضمنها النموذج في توحدها بالصياغة اما نتيجة تقسيم بعضها الى فئات واعطاء قيم 1، 2، ... الخ للفئات او لكون البعض الاخر هي متغيرات هيكلية اصلا جاءت عبر تحويلها من متغيرات نوعية الى متغيرات كمية، فقد تم تحويل المتغير التابع هنا ولاغراض تحليل المرونات فقط الى فئات هي 1-7، 8-15، 15 فاكثر واعطاء القيمة 1 للفئة الاولى و2 للفئة الثانية و القيمة 3 للفئة الثالثة، واعتماد معطيات الجدول رقم (5.6) لهذا الغرض،فنحصل على متوسط جديد للمتغير التابع مقداره (1.27) لاغراض تحليل المرونات.

جدول رقم (5.6)

يوضح المعطيات التي اعتمدت في تغير صياغة المتغير التابع الى فئات بدلا من الصيغة المطلقة لاغراض تحليل المرونات فقط

No. of papers & books published by respondent

		Frequency	Percent	Vaild Percent	Cumulative Percent
Vaild	1.00	2	2.7	2.7	2.7
	2.00	14	18.9	18.9	21.6
	3.00	13	17.6	17.6	39.2
	4.00	3	4.1	4.1	43.2
	5.00	7	9.5	9.5	52.7
	6.00	5	6.8	6.8	59.5
	7.00	2	2.7	2.7	62.2
	8.00	4	5.4	5.4	67.6
	9.00	4	5.4	5.4	73.0

	10.00	2	2.7	2.7	75.7
	11.00	7	9.5	9.5	85.1
	12.00	1	1.4	1.4	86.5
	13.00	2	2.7	2.7	89.2
	14.00	2	2.7	2.7	91.9
	16.00	1	1.4	1.4	93.2
	17.00	3	4.1	4.1	97.3
	18.00	1	1.4	1.4	98.6
	21.00	1	1.4	1.4	100.0
	Total	74	100.0	100.0	

وباستخدام صيغة المرونة اعلاه مع قيم معاملات الانحدار ومتوسط كل من المتغيرات المشمولة بالنموذج المطور مع القيم الجديدة للمتغير التابع، نحصل على درجة مرونة المتغيرات المستقلة التالية :

$$E_{Tit} = 2.234.\frac{0.8243}{1.27} = 1.45$$

$$E_{Nay} = 0.999.\frac{2.067}{1.27} = 1.63$$

$$E_{Age} = 1.324.\frac{2.041}{1.27} = 2.13$$

$$E_{Spe} = 0.357.\frac{3.5811}{1.27} = 1.01$$

$$E_{X14} = 0.458.\frac{2.095}{1.27} = 0.76$$

ومن النتيجة يستدل على ان جميع المتغيرات المشمولة في النموذج باستثناء متغير X_{14} هي بين عالية المرونة (قيمتها اكثرمن 1) او مرنة (قيمتها هي 1) مما يشجع على امكانية توظيفها من حيث المبدأ على لتطوير العملية البحثية.

2-4-6 حالة دراسية رقم C_{6-3}

نماذج لاسلوب وضع سينورياهات وفقا لتحليل حساسية المتغيرات

فلو افترضنا على سبيل المثال ان هناك خيارين من التغيير التي من الممكن ادخلها من قبل احدى الجامعات، ويتمثل هذين الخيارين باحداث زيادة بمقدار 10 ٪ او 50 ٪ قياسا بوضعها الحالي، وتريد الجامعة معرفة مقدار ما سيحدثه كل من هذين الخيارين، ولنتامل المتغيرين (Age) و(X_{14}) وهما باشارة موجبة، اي تخطط الجامعة باتجاه تمديد فترة الابقاء على المتقدمين في العمرمن ذوي الخبرة البحثية بالنسبة للمتغير الاول والسعي مع المؤسسات المعنية بنتائج البحوث على تطبيقها بالنسبة للمتغير الثاني، وكذلك توظيف المتغيرين (Spe) و(Tit) وهما باشارة سالبة من خلال الاهتمام بزيادة الاختصاصات العلمية الصرفة وزيادة عدد التدريسسيين من حملة الالقاب العلمية المتقدمة، والاشارة السالبة قد جاءت كون عملية الترميز قد بدأت بالقيم الاقل مع الاختصاصات العلمية ومع الالقاب العلمية المتقدمة لذلك فهي ظهرت بالاشارة السالبة الصحيحة.

السيناريو الاول (باحداث تطوير بنسبة 10 ٪)

ويتمثل بتوظيف المتغيرات الثلاثة العالية المرونة مع المتغير الرابع المرن واستبعاد متغير E_{x14} غير المرن نحصل على صورة التطوير التالية :

الوضع الحالي وخيارات تطوير الانتاج البحثي	مستوى التغيرالحاصل	نسبة التحسن
الوضع الحالي بمتوسط مقداره	6.72 كتاب وبحث	

التطويرباستخدام 4 متغيرات

- بمقدار 10 ٪	11.33	٪ 69.60
- بمقدار 50 ٪	11.67	٪ 73.66

1.التطوير باستخدام المتغير (Tit)

- الوضع الحالي	0.824	--------
- بمقدار 10٪	0.910	٪10
- بمقدار 50٪	1.241	٪ 50

2.التطوير باستخدام المتغير(Nay)

- الوضع الحالي	2.10	-------
- بمقدار 10٪	2.31	٪ 10
- بمقدار 50٪	3.15	٪ 50

3. التطوير باستخدام المتغير (Age)

- الوضع الحالي	2.04	------
- بمقدار 10 ٪	2.24	٪ 10
- بمقدار 50 ٪	3.06	٪ 50

4. التطوير باستخدام المتغير (Spe)

- الوضع الحالي	3.58	-------
- بمقدار 10 ٪	3.94	٪ 10
- بمقدار 50 ٪	5.01	٪ 50
5.ابقاء الوضع الحالي المتغير (X_{14})	2.10	-------

السيناريو الثاني (احداث تطوير بنسبة 50 ٪)

وعلى افتراض ان السيناريو الثاني هو اشراك المتغيرات الخمسة التي ضمها النموذج، فان عملية التحليل تشير علينا بالنتائج التالية :

الوضع الحالي وخيارات تطوير الانتاج البحثي	مستوى التغيرالحاصل	نسبة التحسن
الوضع الحالي بمتوسط مقداره	6.72 كتاب وبحث	
التطويرباستخدام متغيرات النموذج		
- بمقدار 10 ٪	11.54	٪ 72.00
- بمقدار 50 ٪	12.72	٪ 89.30
1.التطوير باستخدام المتغير (Tit)		
- الوضع الحالي	0.824	--------
- بمقدار 10٪	0.910	٪10
- بمقدار 50٪	1.241	٪ 50
2.التطوير باستخدام المتغير(Nay)		
- الوضع الحالي	2.10	-------
- بمقدار 10٪	2.31	٪ 10
- بمقدار 50٪	3.15	٪ 50
3. التطوير باستخدام المتغير (Age)		
- الوضع الحالي	2.04	------
- بمقدار 10 ٪	2.24	٪ 10
- بمقدار 50 ٪	3.06	٪ 50
4. التطوير باستخدام المتغير (Spe)		
- الوضع الحالي	3.58	-------
- بمقدار 10 ٪	3.94	٪ 10
- بمقدار 50 ٪	5.01	٪ 50
5.ابقاء الوضع الحالي المتغير (X_{14})	2.10	-------
- الوضع الحالي	2.10	--------
- بمقدار 10٪	2.31	٪10
- بمقدار 50٪	3.15	٪ 50

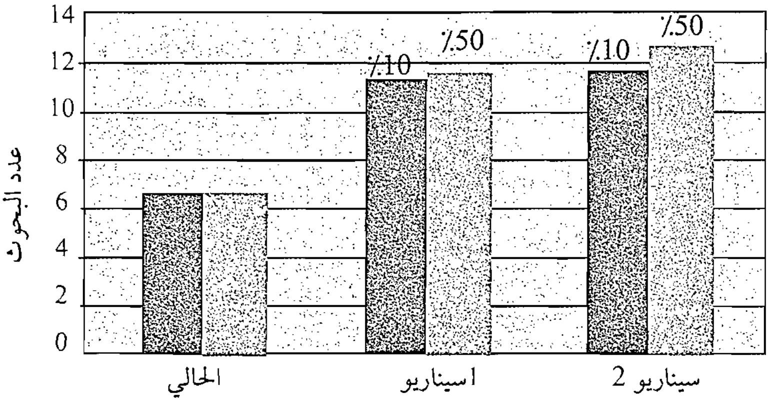
ومن نتائج السيناريو الثاني اعلاه يستدل على امكانية زيادة الانتاج البحثي بنسبة مايقارب 90٪ في حالة حصول ارتفاع في المتغيرات الخمس المستقلة بنسبة 50٪، مقابل زيادة في الانتاج مقداره 72٪ في حال طرأت زيادة مقدارها 10٪ على المتغيرات المستقلة .

اما السيناريو الاول اي في حالة عدم شمول المتغير X_{14} فيشير الى تغير بسيط عن السيناريو الثاني من ناحية اجمالي انتاج البحوث والمؤلفات يتمثل بانخفاض نسبة مقدارها حوالي 1٪ في حالة التطوير بنسبة 10٪ وبنفس النسبة تقرينا ايضا في حالة التطوير بنسبة 50٪.

والشكل البياني (13.6) يعكس صورة التغيرات التي يمكن ان تطرأ على الانتاج البحثي وفقا لكل من السيناريوهات المقترحة مقارنة بالوضع الحالي للانتاج البحثي .

شكل بياني رقم (13.6)

مقارنة بين الانتاج البحثي في ضوء السيناريوهات المقترحة



اما السيناريوهات البديلة الاخرى التي يمكن توفيرها امام متخذ القرار فهي:

- ادخال التطوير على خطوات كأن يكون البدأ بمتغير واحد ومن ثم اثنين وهكذا والتامل بالنتائج او الاعتماد على المتغيرات من نوع Policy variables التي سبق التطرق اليها في الفصل الاول .
- اعتماد نسب تطوير بصورة متدرجة، اي البدا بنسبة 10٪ و15٪ ومن ثم 20٪ وهكذا والتوقف عند نفاذ قدرة المتغيرات على اضافة جديدة للانتاج .

وكما لاحظنا من الحالة الاخيرة في اعلاه نستدل من ان مردود التغيير بمقدار 10٪ على الوضع الحالي سوف يضيف زيادة في الانتاج البحثي بمقدار 72٪، اما خيار رفع نسبة التغيير بنسبة 50٪ فسوف يؤدي الى تحسن اضافي بنسبة مقدارها 17.5٪ عما يضيفه خيار لـ 10٪. وهذا يعود الى درجة المرونة المحدودة التي تتمتع بها بعض المتغيرات ذات التاثير الفعال. ونستدل ايضا من ان التحسن بمقدار 10٪ هو اكثر جدوى اقتصادية من الركون الى خيار لـ 50٪، حيث ان اضافة استثمار اضافي مقداره 40٪ سيقابله مردود مقداره 17.5٪ فقط، وبذلك فان القرار النهائي في حجم التطوير سيتوقف الى حد كبير على الكلفة الحقيقية المطلوبة لكل من الخيارات المطروحة والمدى المطلوب الاستمرار فيه بتحقيق التحسن في العملية البحثية. وقد يكون من المفيد طرح خيارات اخرى قد تتناسب وحدود مرونة المتغيرات الحالية كأن تكون بنسبة 30٪ مثلا او غيرها، او الذهاب ابعد من ذلك من خلال البحث من جديد عن متغيرات جديدة تتميز بمرونات عالية، او استخدام احدى البدائل التي اشرنا اليها هنا بعد فحص نتائجها .

6- 5 الانحدار غير الخطي Non-Linear Regression

وتكون المعادلة هنا على شكل منحني بدلا من خط مستقيم، وذلك اما لكون شكل انتشار المعطيات يشير الى اتجاه الانحناء او بسبب معرفتنا النظرية او نتيجة الخبرة من ان المتغيرات تحت الدراسة علاقاتها غير خطية كما هو الحال مع منحنى الطلب مع وحدة المرونة β التي علاقتها هي :

$$Q = \frac{\beta}{P}$$

حيث ان Q تمثل كمية الطلب و P تمثل سعر البضاعة .

وكما هو الحال مع علاقة معدل الكلفة y والكمية المنتجة x التي تاخذ شكل علاقة الانحدار التربيعي التالي : $y = a + bx + cx^2$

ويتم فيها ايضا استخدام طريقة المربعات الصغرى في ايجاد معاملات الانحدار b's والمعامل الثابت a بعد اضافة بعض الاجراءات الاضافية كزيادة قوة المتغيرات المستقلة، وبالتالي اختلاف حجم معامل كل متغير باختلاف قوة المتغير .

ولايجاد افضل خط انحدار لتضبيط المعطيات، فعادة ما نبدأ بحساب الانحدار الخطي المستقيم لنرى امكانية تخفيض نسبة معنوية من مجموع مربعات البواقي باضافة تربيع الى المتغير المستقل x، ونستمر في اجراء التغيير باضافة التكعيب او اكثر ولغاية الحصول على افضل تضبيط للمعطيات. اما اذا كنا على علم مسبق بطبيعة العلاقة لمتغيرات ظاهرة معينة كأن تكون تربيعية او تكعيبية او اكثر، عندها نبدأ مباشرة باحتساب المعادلة بموجب القوة المطلوبة لـ x .

6- 5- 1 الانحدار غير الخطي البسيط
Simple Non-Linear Regression

ويمكن تلخيص اختلاف الانحدار غير الخطي البسيط عن الخطي البسيط بما يلي :

- ان المعامل الثابت لايظهر بشكل حد مطلق تفصله عن الحد الثاني اشارة + او −
- ان معامل الانحدار ليس مضروبا بالمتغير المستقل x وانما هو على شكل أس power (معادلة أسية)، اي : $y = ax^b$

  او على شكل أساس base كما في حالة دالة القوة، اي : $y = ab^x$
- ان المتغير المستقل x لايظهر بشكله البسيط، وانما على شكل أس او أساس كما لاحظنا في اعلاه، او على شكل لوغاريتم كما في حالة المعادلة النسبية اللوغارتيمية التي شكلها :

$$\frac{y}{x} = a + b\ell inx$$

- ان المتغير التابع y قد لايظهر بشكله الاعتيادي وانما بصيغ اخرى كما في حالة المعادلة النسبية اللوغارتمية اعلاه او باشكال اخرى .

وفي اغلب الحالات يمكن تحويل المعادلات غير الخطية الى معادلات خطية اما باجراء عمليات رياضية كأخذ لوغاريتم طرفي المعادلة او بأعادة تعريف المتغيرات. ففي حالة المعادلة المزدوجة التالية مثلا تصبح معادلة مزدوجة لوغارتيمية، اي :

$$y = ax^b$$

تصبح : $\ell iny = \ell ina + b\ell inx$

حيث نفترض ان :

$$\ell iny = y$$

$$\ell ina = a$$

$$\ell inx = x$$

ج

ويتم ايجاد المعاملات a و b كالاتي :

$$\ell inb = \frac{n\sum \ell inx\ell iny - \sum \ell inx \sum \ell iny}{n\sum\left(\ell inx^2\right) - \left(\sum \ell inx\right)^2}$$

$$\ell ina = \frac{\sum \ell iny - b\sum \ell inx}{n}$$

وهناك حالات يتم تقديرها بمجرد اعادة تعريف المتغيرات ومن دون اجراء عمليات رياضية كما في حالة المعادلة غير الخطية التالية :

$$y = a + \frac{b}{x}$$

$$= a + b\left(\frac{1}{x}\right)$$

وبتعويض $\frac{1}{x}$ بدلا عن x نستطيع تقدير b و a كالاتي :

$$b = \frac{n\sum \frac{1}{x}y - \sum \frac{1}{x}\sum y}{n\sum\left(\frac{1}{x}\right)^2 - \left(\sum \frac{1}{x}\right)^2}$$

$$a = \frac{\sum y - b\sum \frac{1}{x}}{n}$$

مثال (6-2) : أخذت عينة من 9 أسر فكان استهلاكها من البيض y ومعدل دخلها الشهري (بالدينار) x كما مبين في الجدول التالي. والمطلوب تقدير معادلات انحدار التالية لاستهلاك البيض بدلالة الدخل :

- المعادلة الوغارتمية المزدوجة (دالة القوة) $y = ax^b$
- العادلة العكسية $y = a + \frac{b}{x}$
- المعادلة نصف اللوغاريتمية $y = a + b\ell inx$
- تقدير استهلاك البيض لاستهلاك اسرة معدل دخلها الشهري 100 دينار باستخدام كل من العادلات التقديرية اعلاه .

الاسرة	الدخل الشهري (x)	أستهلاك البيض (y)
1	26	28
2	46	53
3	66	70
4	76	90
5	86	91
6	105	115
7	107	130
8	129	142
9	211	190

الحل لـ (6-2) :

- المعادلة الوغارتمية المزدوجة (دالة القوة) $y = ax^b$
- يتم تحويل المعادلة الى معادلة خطية فنحصل على معادلة لوغارتمية مزدوجة هي : $\ell iny = \ell ina + b\ell inx$

- اجراء العمليات الحسابية فيكون لدينا :

$$\sum \ell in x = 39.65$$
$$\sum \ell in x \sum \ell in y = 180.64$$
$$\sum \ell in y = 40.38$$
$$\sum \left(\ell in x^2\right) = 177.67$$

$$\ell in b = \frac{n\sum \ell in x \ell in y - \sum \ell in x \sum \ell in y}{n\sum \left(\ell in x^2\right) - \left(\sum \ell in x\right)^2}$$

$$= \frac{9(180.64) - (39.65)(40.38)}{9(177.57) - (39.65)} = \frac{24.69}{26.91} = 0.92$$

$$\ell in a = \frac{\sum \ell in y - b\sum \ell in x}{n}$$

$$= \frac{40.38 - (0.92)(39.65)}{9} = 0.43$$

اي : $a = e^{0.43} = 1.54$

$$y = 1.54x^{0.92}$$

او $\ell in y = 0.43 + 0.92 \ell in x$

عند x = 100 فالمتوقع ان يكون الاستهلاك هو :

$$\ell in \hat{y} = 0.43 + 0.92 \ell in(100)$$
$$= 0.43 + (0.92)(4.605)$$
$$= e^{4.67}$$
$$\hat{y} = 106.7$$

- المعادلة العكسية $y = a + \frac{b}{x}$

- لدينا :

$$\sum \frac{1}{x} y = 90.773$$

$$\sum y = 909$$

$$\sum \frac{1}{x} = 0.13$$

$$\sum \left(\frac{1}{x}\right)^2 = 0.00267$$

- نجد قيم كل من a , b :

$$b = \frac{n\sum \frac{1}{x} y - \sum \frac{1}{x} \sum y}{n\sum \left(\frac{1}{x}\right)^2 - \left(\sum \frac{1}{x}\right)^2}$$

$$= \frac{9(9.773) - (0.13)(909)}{9(0.00267) - (0.13)} = -4156.9$$

$$a = \frac{\sum y - b\sum \frac{1}{x}}{n}$$

$$= \frac{909 - (-4156.9)(0.13)}{9} = 161.1$$

$$y = 161.1 - \frac{4156.9}{x}$$

وعند $x = 100$ نحصل على :

$$\hat{y} = 161.1 - \frac{4156.9}{100} = 119.5$$

- المعادلة نصف اللوغاريتمية $y = a + b\ell inx$

- لدينا :

$$\sum y\ell inx = 4235.5$$

$$\sum \ell inx = 39.65$$

$$\sum y = 909$$

$$\sum (\ell inx)^2 = 177.67$$

- نجد قيم كل من a , b :

$$b = \frac{n\sum y\ell inx - \sum \ell inx \sum y}{n\left(\sum \ell inx^2\right) - \left(\sum \ell inx\right)^2}$$

$$= \frac{9(4235.5) - (39.65)(909)}{9(177.67) - (39.65)^2} = \frac{2077.65}{26.907} = 77.216$$

$$a = \frac{\sum y - b\sum \ell inx}{n}$$

$$= \frac{909 - (77.216)(39.67)}{9} = -239.15$$

$$\hat{y} = -239.35 + 77.216\ell inx$$

- والتعويض عن x بالقيمة 100 نحصل على :

$$\hat{y} = -239.35 + 77.216\ell in100 = 116.3$$

ولاجل اختيار التقدير الافضل من بين نتائج المعادلات اعلاه، يتم استخراج مجموع مربعات الانحرافات لكل من المعادلات اعلاه، اي: $y - \hat{y}$، ومن ثم تربيع الانحرافات وجمعها فتلك التي تعطي اصغر مجموع للمربعات يتم اخيارها كأفضل معادلة للتقدير. وفي بعض الحالات يدلنا شكل انتشار المعطيات المبين نموذجه في الشكل البياني رقم (14.6) التالي الى متوسطات $\mu_{y/x}$ يمكن تضبيطها (fitting it) من خلال تمثيلها بمنحني أسي Exponential Regression Curve والذي صيغة معادلته في حالة العينة هي :

$$\mu_{y/x} = cd^x$$

ولنرمز لـ $\mu_{y/x}$ بـ $\hat{y}$، واخذ لوغاريتم الاساس 10 نحصل على معادلة الانحدار الخطي التالية :

$$\log y = \log c + \log d$$

باحلال :

$$a = \log c$$

$$b = \log d$$

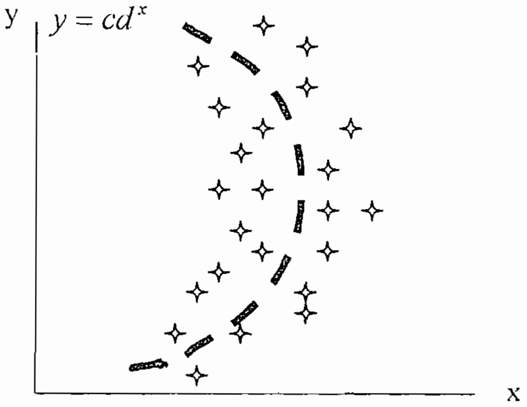
نحصل على :

$$\log y = a + bx$$

فيصبح بالامكان ايجاد قيم المعاملات a, b باعتماد صيغ الانحدار الخطي ومن ثم تحديد قيم d, c باخذ اللوغاريتم المقابل antilog وكما مبين في المثال (6-3) التالي .

شكل بياني رقم (14.6)

يوضح نموذج منحني معادلة الانحدار الاسية



مثال (6-3): المعطيات التالية تمثل عدد الطلاب المسجلين في احدى المدارس الابتدائية خلال السنوات السبع الاخيرة، والمطلوب ايجاد المعادلة الاسية للتنوء بعدد الطلاب المتوقع تسجيلهم بعد 6 سنوات .

السنة (x)	1	2	3	4	5	6	7
عدد الطلاب (y)	304	341	393	493	548	670	882

الحل لـ (3-6) :

- نستخرج قيم y وهي :

2.945 , 2.826 , 2.739 , 2.66 , 2.594 , 2.533 , 2.483

- اجراء العمليات الحسابية لتقدير المعاملات a , b وكالاتي :

$$\sum x = 28$$
$$\sum \log y = 18.78$$
$$\sum x^2 = 140$$
$$\sum x \log y = 77.237$$

$$b = \frac{n\sum \log y - \left(\sum x\right)\left(\sum \log y\right)}{n\sum x^2 - \left(\sum x\right)^2}$$
$$= \frac{(7)(77.237) - (28)(18.78)}{(7)(140) - (28)^2} = 0.076$$

$$a = \frac{\sum \log y - b\sum x}{n}$$
$$= \frac{18.78 - (0.076)(28)}{7} = 2.379$$

- وعليه فان :

$$c = 10^{2.379} = 239$$
$$d = 10^{0.076} = 1.19$$

- وبالتعويض يكون لدينا :

$$y = cd^x = (239)(1.19)^x$$

- عدد الطلاب المتوقع تسجيلهم بعد 6 سنوات، اي $x = 13$ هو :

$$\hat{y} = (239)(1.19)^{13}$$
$$= (239)(13.589) = 2167$$

6-5-2 حالة دراسية C_{6-3}

استخدام برنامج SPSS في تحليل الانحدار غير الخطي البسيط

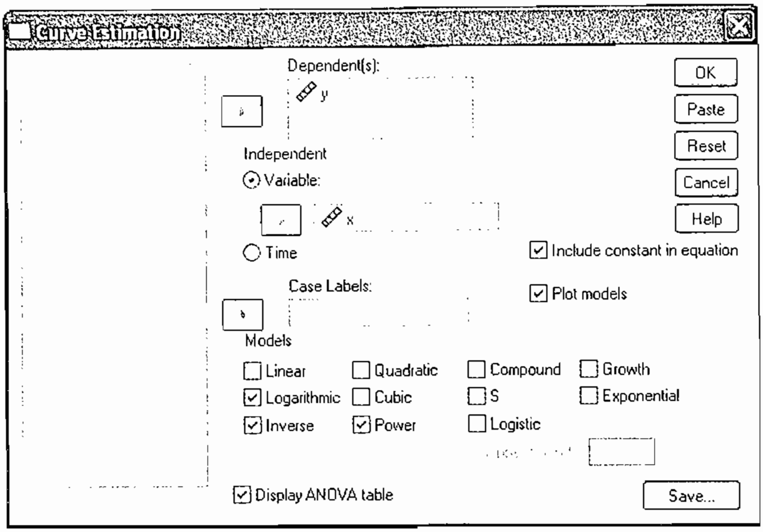
استخدم معطيات المثال (3.6) اعلاه التالية، في ايجاد كل من المعادلة الاسية Power واللوغارتمية Logarithmic والعكسية Inverse .

السنة (x)	1	2	3	4	5	6	7
عدد الطلاب (y)	304	341	393	493	548	670	882

- الدخول الى البرنامج وانشاء ملف بالمعطيات اعلاه، ويتم استدعاء القائمة Analysis ومنها الامر الفرعي Regression ومن ثم الخيار Curve Estimation، فيظهر لنا مربع الحوار Curve Estimation المبن في الشكل البياني رقم (15.6). وفي مربع الحوار يتم استخدام السهم الجانبي لنقل المتغير التابع y الى تحت Dependent واستخدام السهم الجانبي الثاني لنقل المتغير المستقل x الى تحت Independent، بعد ان يتم التاشير عند Variable.
- وعند نفس مربع الحوار Curve Estimation، يتم ايضا التاشير تحت Models عند النماذج المطلوبة وهي Logarithmic و Inverse و Power، ومن ثم التاشير في نهاية المربع عند Table Display AOVA للحصول على اشكال المنحنيات الناتجة عن تضبيط النماذج الثلاثة .
- الكبس على ايقونة Ok لنحصل على مخرجات التحليل المبينة في الشكل البياني رقم (16.6) و الجداول رقم (6.6) .

الشكل البياني رقم (15.6)

لوحة حوار النماذج غير الخطية

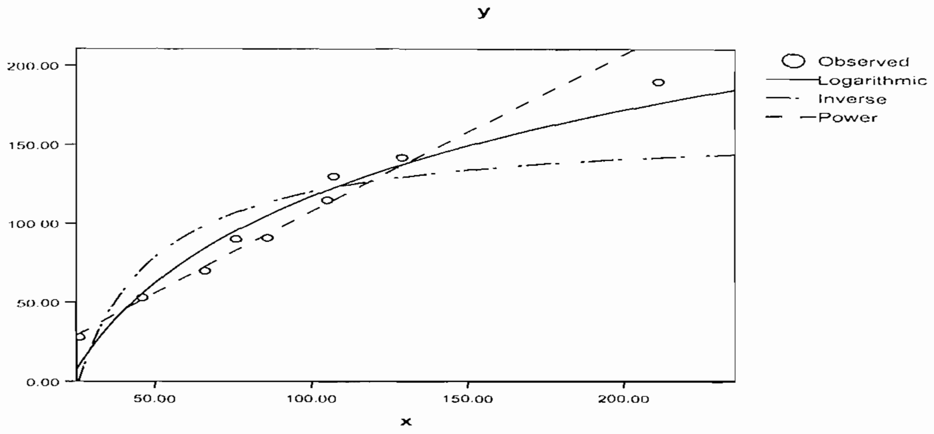


وعند التمعن في الشكل البياني والجدول التالي، نستدل بوضوح من ان النموذج الاسي Power كان الاكثر تضبيطا للمعطيات، يليه النموذج اللوغارتيمي Logarithmic، وان نتائج التحليل المبينة في جداول المخرجات جاءت تعزيزا لذلك الاستنتاج وكما يتضح من المقارنة البسيطة التالية :

المعيار	النموذج الاسي **Power model**	الموذج اللوغارتيمي **model Logarithmic**	النموذج العكسي **Inverse model**
R	0.990	0.973	0.856
R^2	0.981	0.946	0.73
F	360.99 Sig at 0.000	122.613 Sig at 0.000	19.198 Sig at 0.003
Beta	0.990	0.973	-0.856
t Sig at	0.000	0.000	0.003

الشكل البياني رقم (16.6)

مقارنة القيم الحقيقية مع نتائج النماذج غير الخطية وهي : الاسي، اللوغارتيمي والعكسي



جداول رقم (6.6)

مخرجات تحليل الانحدار غير الخطي البسيط

Model Description :Curve Fit

Model Name		MOD_2
Dependent Variable	1	y
Equation	1	Logarithmic
	2	Inverse
	3	Power(a)
Independent Variable		x
Constant		Included
Variable Whose Values Label Observations in Plots		Unspecified

a The model requires all non-missing values to be positive.

Variable Processing Summary

		Variables	
		Dependent	Independent
		y	x
Number of Positive Values		9	9
Number of Zeros		0	0
Number of Negative Values		0	0
Number of Missing Values	User-Missing	0	0
	System-Missing	0	0

Model Summary :Logarithmic

R	R Square	Adjusted R Square	Std. Error of the Estimate
.973	.946	.938	12.251

The independent variable is x.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	18403.346	1	18403.346	122.613	.000
Residual	1050.654	7	150.093		
Total	19454.000	8			

The independent variable is x.

Coefficients

	Unstandardized Coefficients		Standardize Coefficients		
	B	Std. Error	Beta	t	Sig.
ln(x)	79.134	7.147	.973	11.073	.000
Constant	-247.190	31.709		-7.796	.000

Model Summary :Inverse

R	R Square	Adjusted R Square	Std. Error of the Estimate
.856	.733	.695	27.250

The independent variable is x.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	14256.036	1	14256.036	19.198	.003
Residual	5197.964	7	742.566		
Total	19454.000	8			

The independent variable is x.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
1 / x	-4146.738	946.400	-.856	-4.382	.003
Constant	161.588	16.544		9.767	.000

Model Summary : Power

R	R Square	Adjusted R Square	Std. Error of the Estimate
.990	.981	.978	.085

The independent variable is x.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	2.616	1	2.616	360.989	.000
Residual	.051	7	.007		
Total	2.667	8			

The independent variable is x.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
ln(x)	.944	.050	.990	19.000	.000
(Constant)	1.398	.308		4.538	.003

The dependent variable is ln(y).

6- 5- 3 الانحدار غير الخطي المتعدد

Non-Linear Multiple Regression

(1) معادلة الانحدار التربيعية

Quadratic Regression Equations

وتعتبر من ابسط انواع الانحدار غير الخطي وتدعى ايضا بالانحدار من الدرجة الثانية، وتتم باضافة العنصر x^2 الى معادلة الانحدار الخطي البسيط لنحصل على معادلة الانحدار التربيعي التي تاخذ صيغة العلاقة التالية :

$$y = a + bx + cx^2$$

ويكون شكل المنحني على صيغة أجزاء او مقاطع عمودية متكافئة ويدعى بمنحنى الاجزاء المتكافئة Parabola وكما مبين في الاشكال البيانية (17.6) و(18.6)، ويكون مفتوحا الى الاعلى عندما يكون المعامل c موجبا، وان أوطئ نقطة في المنحنى تدعى بقمة الرأس Vertex ويتحدد موقعها على المحور الافقى بالصيغة التالية :

$$x = \frac{-b}{2c}$$

اما عندما يكون المعامل c- سالبا فان فتحة منحنى الاجزاء المتكافئة تكون باتجاه الاسفل. ولاجل تحويل معادلة الانحدار التربيعية الى معادلة انحدار خطية نفترض بان :

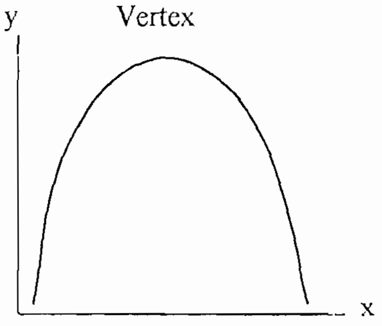
$$b_1 = b$$
$$x_1 = x$$
$$b_2 = c$$
$$x_2 = x^2$$

فتحصل على الصيغة التالية :

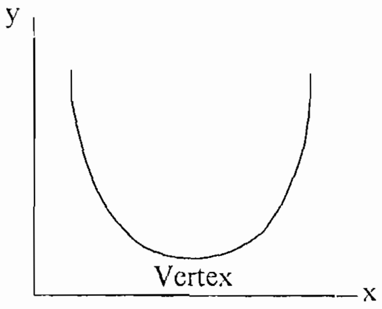
$$y = a + b_1x_1 + b_2x_2$$

وبذلك يمكننا استخدام طريقة المربعات الصغرى لايجاد قيم كل من a , b_1 , b_2 كما هو الحال في معادلة الانحدار الخطي المتعدد، وتصبح عملية حساب معامل الارتباط R وكذلك الخطأ المعياري S_e بنفس الصيغة للانحدار الخطي المتعدد.

شكل بياني رقم (18.6) يوضح شكل منحنى المعادلة التربيعية في حالة المعامل c سالباً



شكل بياني رقم (17.6) يوضح شكل منحنى المعادلة التربيعية في حالة المعامل c موجباً



مثال (6-4): المعطيات التالية تخص عدد الوحدات المنتجة من بطاريات السيارات (بالاف) من قبل احد المصانع x، ومعدل كلفة الوحدة المنتجة (بالدولار) y، والمطلوب ايجاد معادلة الانحدار التربيعي مع تقدير كلفة الوحدة عند انتاج $x = 2.5$.

عدد الوحدات امنتجة (x)	1	2	3	4	5
معدل كلفة الوحدة (y)	6	3	2	3	5

الحل لـ (6-4) :

- نجري العمليات الحسابية لمتطلبات ايجاد معادلة الانحدار التربيعي وكالاتي:

y^2	x_2^2	x_1^2	x_1x_2	x_2y	X_1y	$x^2=x_2$	y	$x=x_1$
36	1	1	1	6	6	1	6	1
9	16	4	8	12	6	4	3	2
4	81	9	27	18	6	9	2	3
9	256	16	64	48	12	16	3	4
25	625	25	125	125	25	25	5	5

- نستخدم صيغ الانحدار الخطي المتعدد في ايجاد قيم معاملات b_1, b_2, a فنحصل على :

$$b_1 = \frac{\left[n\sum x_1y - \sum x_1\sum y\right]\left[n\sum x_2^2 - \left(\sum x_2\right)^2\right] - \left[n\sum x_2y - \sum x_2\sum y\right]\left[n\sum x_1x_2 - \sum x_1\sum x_2\right]}{\left[n\sum x_1^2 - \left(\sum x_1\right)^2\right]\left[n\sum x_2^2 - \left(\sum x_2\right)^2\right] - \left[n\sum x_1x_2 - \sum x_1\sum x_2\right]^2}$$
$$= -0.5343$$

$$b_2 = \frac{\left[n\sum x_2y - \sum x_2\sum y\right]\left[n\sum x_1^2 - \left(\sum x_1\right)^2\right] - \left[n\sum x_1y - \sum x_1\sum y\right]\left[n\sum x_1x_2 - \sum x_1x_2\right]}{\left[n\sum x_1^2 - \left(\sum x_1\right)^2\right]\left[n\sum x_2^2 - \left(\sum x_2\right)^2\right] - \left[n\sum x_1x_2 - \sum x_1\sum x_2\right]^2}$$
$$= 0.857$$

$$a = \frac{\sum y - b_1\sum x_1 - b_2\sum x_2}{n}$$
$$= 10.4$$

- وبتطبيق صيغة معادلة الانحدار الخطي نحصل على :

$$\hat{y} = a + b_1x_1 + b_2x_2$$

$$= 10.4 - 5.343x_1 + 0.857x_2$$

- وباستخدام المتغيرات الاصلية نحصل على معادلة الانحدار التربيعي وهي :

$$\hat{y} = 10.4 - 5.343x + 0.856x^2$$

- وبتعويض $x = 2.5$ فان معدل كلفة انتاج الوحدات المتوقعة هو :

$$\hat{y} = 10.4 - 5.343(2.5) + 0.856(2.5)^2 = 2.4$$

- وان قيمة نقطة قمة رأس المنحنى X هي :

$$\text{vertex} - \text{value}, x = -\frac{b}{2c} = -\frac{(-5.343)}{2(0.857)} = 3.117$$

(2) معادلة الانحدار التكعيبي Cubic Regression Equation

وتدعى ايضا بمعادلة الانحدار غير الخطية من الدرجة الثالثة، وهي امتداد لمعادلة الانحدار التربيعية، ويتم استخدامها عندما تتطلب المعطيات اللجوء الى اضافة القوة 3 الى معادلة تضبيط المعطيات، ويصبح شكل العلاقة كالاتي :

$$y = a + bx + cx^2 + dx^3$$

ويتم ايجادها من خلال تحويلها الى معادلة انحدار خطية وذلك بتغيير المعاملات والمتغيرات على الوجه الاتي :

$$b_1 = b$$
$$x_1 = x$$
$$b_2 = c$$
$$x_2 = x^2$$
$$b_3 = d$$
$$x_3 = x^3$$

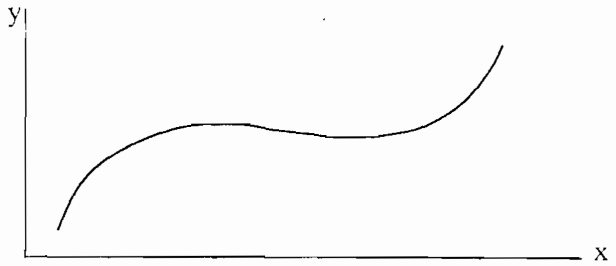
فتصبح معادلة انحدار خطية، اي :

$$y = a + b_1x_1 + b_2x_2 + b_3x_3$$

وباخذ منحني معادلة الانحدار التكعيبي الشكل رقم (19.6) وهو شكل مقعر Concave يكون جزءه الايمن مقعر الى الاعلى، اما جزءه الايسر فيكون اتجاه تقعره نحو الاسفل.

شكل بياني رقم (19.6)

يوضح نموذج لمنحنى معادلة الانحدار التكعيبي



ونتابع في الحالة الدراسية C_{6-4} التالية كيفية الاستعانة بالحاسوب من خلال استخدام برنامج SPSS في تحليل الانحدار غير الخطي المتعدد .

6-5-4 حالة دراسية رقم C_{6-4}

استخدام برنامج SPSS في تحليل الانحدار غير الخطي المتعدد

نستخدم معطيات المثال اعلاه (4.6) التالية التي تخص عدد الوحدات المنتجة من بطاريات السيارات (بالاف) من قبل احد المصانع x، ومعدل كلفة الوحدة المنتجة (بالدولار) y ، والمطلوب ايجاد معادلتي الانحدار التربيعي والتكعيبي.

عدد الوحدات امنتجة(x)	1	2	3	4	5
معدل كلفة الوحدة(y)	6	3	2	3	5

لاتختلف الاجراءات المطلوبة لاستخدام برنامج SPSS في تحليل الانحدار غير الخطي المتعدد عن تلك التي تم ذكرها مع تحليل الانحدار غير الخطي البسيط اعلاه، باستثناء ان يكون التاشير هنا عند Quadratic او Cubic او كلاهما الموجودة تحت عنوان Models في مربع الحوار المبين في الشكل رقم (20.6) .

فبانشاء ملف معطيات المثال (4.6)، واخضاعه للتحليل وفقا للاجراءات المنوه عنها، نحصل على المخرجات المبينة في الشكل البياني رقم (21.6) وفي الجداول رقم (7.6).

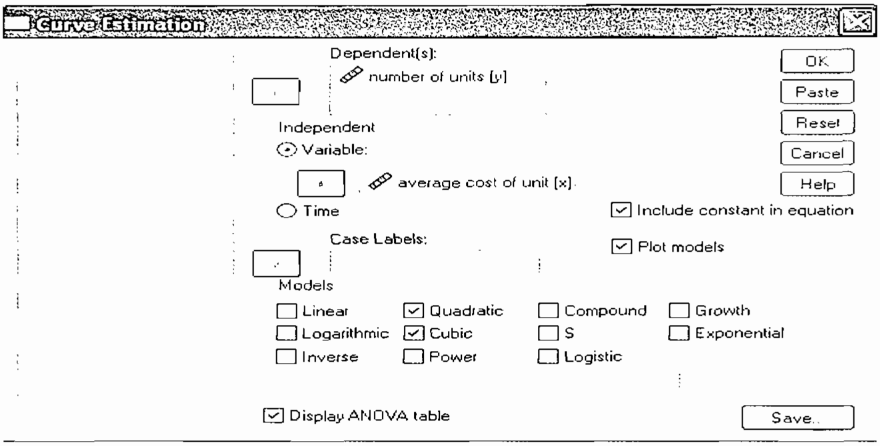
ومن نتائج مخرجات التحليل، وكما يتضح جليا من الشكل البياني رقم (21.6) بان نماذج التحليل غير الخطي المتعدد هو المناسب لتضبيط المعطيات، وخاصة النموذح التربيعي Quadratic Model حيث جاءت كل من المعطيات الحقيقية والتقديرية شبه متطابقة. وهذا ما يفسر معاملي التحديد R^2 هي 0.99 و 0.989 لنموذجي التربيعي والتكعبي على التوالي.

اما اختبار F على نطاق النموذج و t على مستوى المتغيرات فهي معنوية عند $\alpha = 0.05$.

الشكل البياني رقم (20.6)

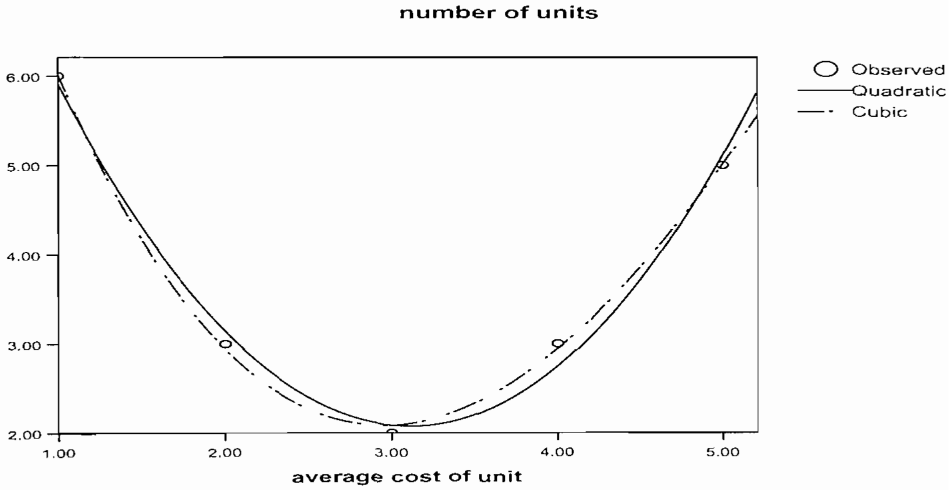
مربع حوار Regression : Curve Estimation

في تحليل الانحدار غير الخطي المتعدد



الشكل البياني رقم (21.6)

يوضح نموذجي التربيعي والتكعيبي في تضبيط معطيات المثال رقم (4.6)



جداول رقم (7.6)

مخرجات تحليل الانحدار غير الخطي المتعدد لمعطيات المثال (4.6)

Model Description:Curve Fit

Model Name		MOD_3
Dependent Variable	1	number of units
Equation	1	Quadratic
	2	Cubic
Independent Variable		average cost of unit
Constant		Included
Variable Whose Values Label Observations in Plots		Unspecified
Tolerance for Entering Terms in Equations		.0001

Variable Processing Summary

		Variables	
		Dependent	Independent
		number of units	average cost of unit
	Number of Positive Values	5	5
	Number of Zeros	0	0
	Number of Negative Values	0	0
Number of Missing Values	User-Missing	0	0
	System-Missing	0	0

Model Summary : Quadratic

R	R Square	Adjusted R Square	Std. Error of the Estimate
.995	.989	.979	.239

The independent variable is average cost of unit.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	10.686	2	5.343	93.500	.011
Residual	.114	2	.057		
Total	10.800	4			

The independent variable is average cost of unit.

Coefficients

	Unstandardized Coefficients		Standardize Coefficients	t	Sig.
	B	Std. Error	Beta		
average cost of unit	-5.343	.391	-5.141	-13.675	.005
average cost of unit ** 2	.857	.064	5.044	13.416	.006
(Constant)	10.400	.513		20.285	.002

Model Summary :Cubic

R	R Square	Adjusted R Square	Std. Error of the Estimate
.999	.999	.995	.120

The independent variable is average cost of unit.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	10.786	3	3.595	251.667	.046
Residual	.014	1	.014		
Total	10.800	4			

The independent variable is average cost of unit.

Coefficients

	Unstandardized Coefficients		Standardize Coefficients	t	Sig.
	B	Std. Error	Beta		
average cost of unit	-7.310	.769	-7.034	-9.511	.067
average cost of unit ** 2	1.607	.285	9.458	5.634	.112
average cost of unit ** 3	-.083	.031	-2.585	-2.646	.230
(Constant)	11.800	.588		20.069	.032

6- 6 تحليل المركبات الاساسية
Principal Component Analysis

وهو الموضوع الذي تم التطرق اليه في الفقرة (4.5) من الفصل الخامس الذي تم فيه عرض حالة دراسية باستخدام برنامج SPSS، كما سيتم استخدام الطريقة في الفصل الاحق بدمجها مع طريقة تحليل الانحدار للخروج بمنهجية جديدة في بناء النماذج.

6-7 تحليل دالة التمييز Discriminent Function Analysis

6-7-1 خصائص واستخدامات دالة التمييز

تستخدم بشكل واسع في عملية تصنيف بيانات (معطيات) المتغير التابع حسب طبقات او مواصفات محددة متمثلة بالمتغيرات المستقلة، ويمكن اجمال اهم اهداف استخدام الدالة المميزة بما يلي :

- البحث عن الفروقات بين المجاميع
- تحديد اهم العوامل التي تميز بين المجاميع
- استبعاد المتغيرات التي تاثيرها يكون ضعيفا في دالة التمييز
- تصنيف االظاهرة او الظواهر الى مجاميع محددة
- لاختبار ان كان التصنيف وفقا لما هو متوقع ام لا

وبذلك فان تحليل التمييز يستخدم لبناء توقعات للمتغير التابع الذي يتكون من مجموعتين (صنفين) بالاعتماد على خصائص الظاهرة المدروسة من خلال توليد دالة مميزة لمجموعة من المجموعتين التي يتكون منها المتغير التابع، حيث تقوم الدالة بايجاد معاملات المتغيرات التي تستخدم في عملية التمييز للمجموعة المعنية. ويتم ذلك بالاعتماد على تحليل الانحدار، فاذا رمزنا للمتغير التابع في الدالة المييزة Z فيصبح شكل النموذج كالاتي :

$$Z = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots . + \beta_k X_{ik} + \varepsilon_i$$

حيث ان :

X_i = المتغيرات المستقلة

ε_i = متغير الاخطاء العشوائية (البواقي)

α = المعامل الثابت ويتم تقديره باستخدام طريقة المربعات الصغرى ،

اما فرضيات التحليل ومعايير الجودة فهي ذاتها التي تم التطرق اليها بتفصيل في طريقة تحليل الانحدار.

كما بالامكان بناء نموذج لاكثر من مجموعتين او صنفين بالاعتماد على العلاقات الخطية للمتغيرات المستقلة للتمييز بين المجموعات. ففي حالة المجموعات التي تعود لنوع متجانس في الصفة، يتم ذلك باختيار الامر الفرعي k-Means Cluster من الامر الرئيسي Analysis لبرنامج SPSS. اما في حالة التمييز بين اكثر من مجموعتين ولكن غير متجانسة في الصفة فيتم استخدام الامر الفرعي Hierachical Cluster. مع الاشارة الى ان قيم المجموعات يجب ان تكون اعداد صحيحة Integer مع ضرورة تحديد الحد الادنى والاعلى لهذه القيم.

6-7-2 حالة دراسية $C_{6\text{-}6}$

استخدام برنامج SPSS في تحليل دالة التمييز

سيتم استخدام معطيات دراسة مستوى الرضا والعوامل المؤثرة على الانتاج البحثي، واختيار متغير الشهادة Deg الذي يتكون من مجموعتين هم حاملي الدكتوراه وقيمتها 1 وحاملي الماجستير وقيمتها 2 للتمييز في انتاجهم البحثي وفقا للمتغيرات المستقلة التي تم شرحها في فصل التحليل الوصفي والبالغ عددها 25 متغيرا، وتوظيف طريقة الخطوات Stepwise في اختيار طاقم المتغيرات،

والنموذج الذي سيتم تطويره هو لتقييم مدى تاثير مجموعة المتغيرات المستقلة على الباحثين لكل من حاملي شهادة الدكتوراه واؤلئك من حملة شهادة الماجستير، حيث تم تقسيم عينة المعطيات الى (35) مشاهدة من حملة الدكتوراه و (39) مشاهدة لحملة شهادة الماجستير. ويتم اخضاع المعطيات لدالة تحليل التمييز Discriminant Analysis، والدالة تفترض التجانس بين المجتمعين (حملة الدكتوراه ومجتمع حملة الماجستير)، و الهدف هو التوصل

الى تحديد المتغيرات التي لها تاثير معنوي في التمييز بين المجتمعين الخاضعين للتحليل. وتتمثل هذه المتغيرات بخصائص الباحثين (الجنس، العمر، بلد الحصول على اخر شهادة، الاختصاص، الدخل، ... الخ) ومتغيرات مستوى الرضا عن الظروف المحيطة بالعملية البحثية التي يقومون بها.

ان معايير نجاح عملية التميز التي تفرزها عملية التحليل على مجمل المعادلة تتمثل بما يلي :

- نسبة العدد من كل من المجتمعين الذين تستطيع الدالة تصنيفهم بصورة صحيحة
- معامل الارتباط بين المتغيرات التابعة **R** , canonical correlation coefficient
- اما على نطاق المتغيرات، فان معايير النجاح تعتمد على مقاييس كل من: مقياس Wilks lambda، مقياس F-ratio

تتلخص خطوات استخدام برنامج SPSS في عملية التحليل بما يلي :

- استدعاء القائمة Analysis ومنها الامر الفرعي Classify ومن ثم الخيار Discriminant فيظهر لنا مربع الحوار **Discriminant Analysis** المبين في الشكل البياني رقم (22.6)، وفيه يتم استخدام السهم الجانبي لنقل المتغير التابع Deg الى خانة Grouping Variable بعدها الكبس على ايقونة Define Range الموجودة اسفلها لتظهر لنا لوحة صغيرة بذات الموقع يتم فيها تدوين 1 في خانة Manim و2 في خانة Maxim، ونقل المتغيرات المستقلة التي تستخدم للتمييز بين مجموعتي الدكتوراه 1 ومجموعة الماجستير 2 الى خانة Independents، ومن ثم التاشير عند Use Stepwise Method،

- الكبس على ايقونة Statistics فتظهر لوحة Discriminant Analysis : Statistics المبينة في الشكل رقم (23.6)، يتم عليها التاشير عند المؤشرات المطلوب استخدامها وظهورها، ثم الكبس على ايقونة Continue للعودة الى مربع الحوار،
- الكبس على ايقونة Method فتظهر لوحة Discriminant Analysis : Stepwise Method المبينة في الشكل البياني رقم (24.6)، ويتم فيها تحديد مستوى المعنوية المطلوب بموجبها شمول المتغير في النموذج، او الابقاء على مستوى المعنوية المحدد مسبقا وذلك تحت عنوان Criteria، والتاشير عند Wilk's Lambda و F-ratio تحت عنوان Method، وفي حالة الرغبة عند Summary of Steps تحت عنوان Display بعدها الكبس على Continue للعودة الى مربع الحوار من جديد .
- الكبس على ايقونة Classify فتظهر لوحة Discriminant Analysis :Classification المبينة في الشكل البياني رقم (25.6)، وفيها يتم التاشير على All group equal تحت عنوان Prior Probabilities وعلى الاشكال البيانية المطلوبة تحت العنوان Plots ، ومن ثم الكبس على ايقونة Continue للعودة الى مربع الحوار ،
- الكبس على ايقونة Ok للحصول على المخرجات المبينة في الجداول رقم (8.6)،

ومن مجموعة الجداول رقم (8.6) نستدل على :

1. ان مجموعة المتغيرات التي تضمنتها النماذج استطاعت تفسير مانسبته 71.4 ٪ من التباين (R^2)، وان نسبة المشاهدات التي لم يتم تصنيفها لايتجاوز 8.1 ٪ من المشاهدات فقط

2. وان قيمة معامل الارتباط canonical correlation coefficient, R هو

R = 0.845 ، عالي المعنوية، اي عند Sig. at 0.000

3. تضمن النموذج 4 متغيرات من مجموع 25 متغيرا وهذه المتغيرات المعنوية التي تضمنها نموذج دالة التمييز وحسب اهميتها هي : العمر (Age)، اللقب الاكاديمي (Tit)، الجامعة التي يعمل فيها (Uni)، الراتب الشهري (Spm).

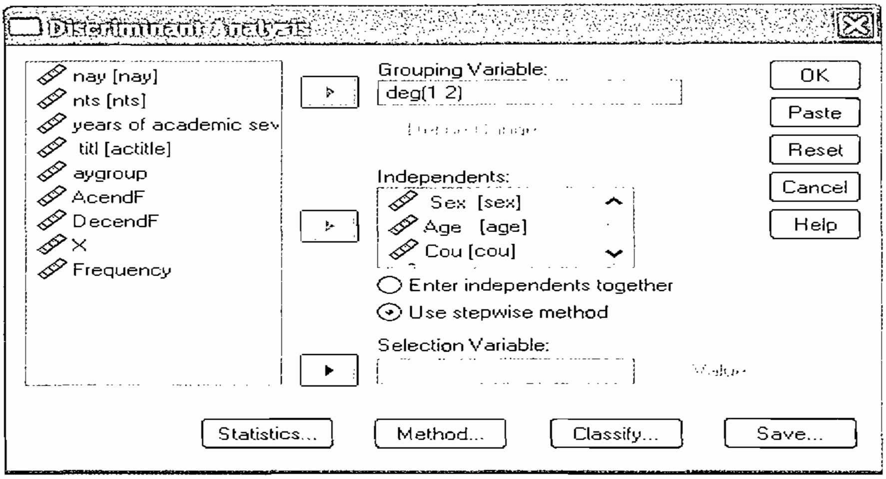
4. ان كل من مقياسي Wilks ' lambda and F جاءت بمعنوية عالية لجميع المتغيرات التي تضمنها النموذج Sig at 0.000 ،

وان شكل النموذج الذي تم تطويره هو :

$$Z = -7.187(\text{constant}) + 0.832\,\text{age} + 2.098\,\text{tit} + 0.508\,\text{uni} - 0.001\,\text{spm}$$

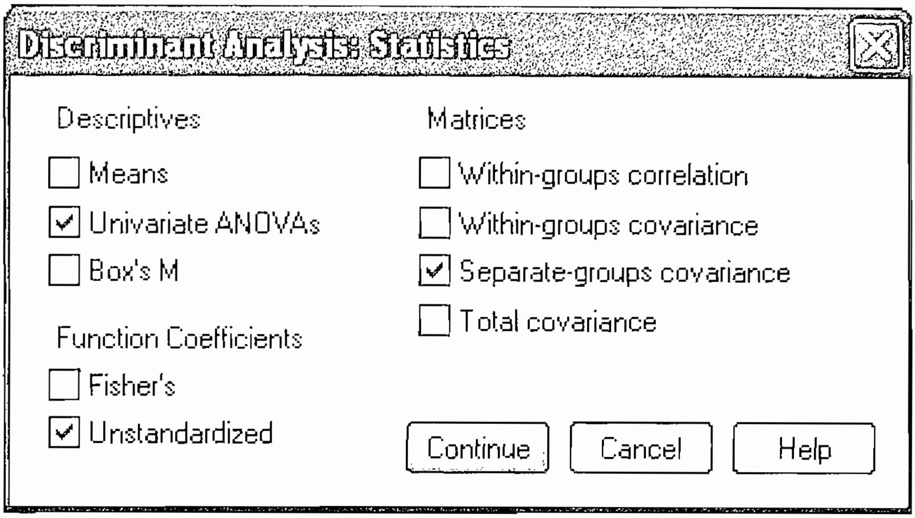
الشكل البياني رقم (22.6)

مربع حوار Discriminant Analysis



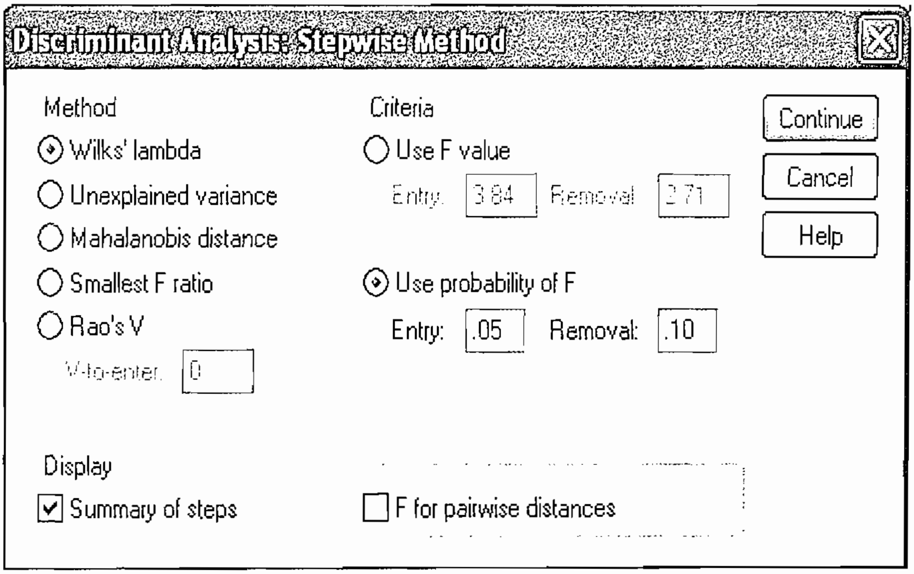
الشكل البياني رقم (23.6)

Discriminant Analysis : Statistics لوحة



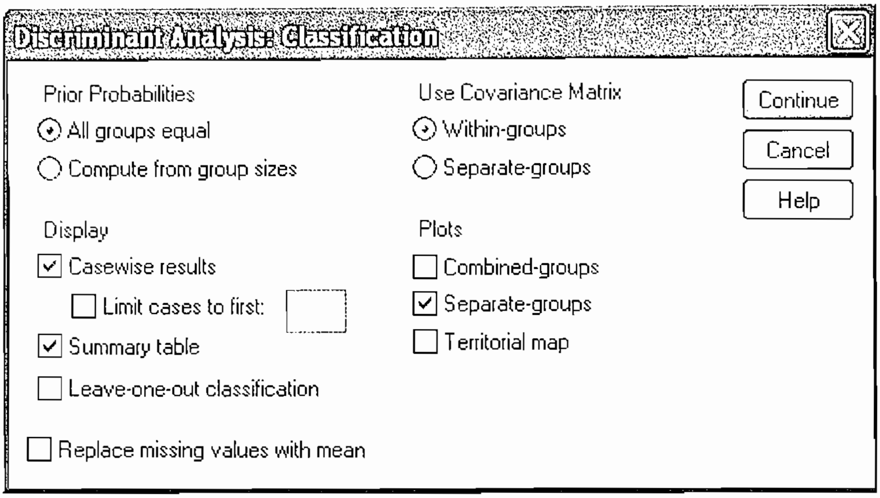
الشكل البياني رقم (24.6)

Discriminant Analysis : Stepwise Method لوحة



الشكل البياني رقم (25.6)

لوحة Discriminant Analysis : Classification



مجموعة جداول رقم (8.6)

تمثل مخرجات التحليل باستخدام الدالة المميزة Discriminant Analysis

Classification Results

			Predicted Group Membership		
		Deg	1.00	2.00	**Total**
Original	**Count**	1.00	34	34	39
		2.00	5	34	39
	%	1.00	97.1	2.9	100.0
		2.00	12.8	87.2	100.0

* 91.9% of original grouped cases correctly classified

Variables Entered/ Removed[1234]

Step	Entered	Wilk's Lambda							
						Exact F			
		Statistic	df1	df2	df3	statistic	df1	df2	df3
1	Titl	.440	1	1	72.000	91.571	1	72.000	.000
2	Age	.359	2	1	72.000	63.494	2	71.000	.000
3	Spm	.320	3	1	72.000	49.662	3	70.000	.000
4	Uni	.286	4	1	72.000	43.104	4	69.000	.000

Summary of Canonical Discriminate Functions

Wilk's Lambda

Step	No. of Variables	Lambda	df1	df2	df3	Exact F			
						Statistic	Df1	Df2	Sig.
1	1	.440	1	1	72	91.571	1	72.000	0.000
2	2	.359	2	1	72	63.494	2	71.000	.0000
3	3	.320	3	1	72	49.662	3	70.000	0.000
4	4	.286	4	1	72	43.104	4	69.000	0.000

Wilk's Lambda

Test of Function (s)	Wilk's Lambda	Chi-square	df	Sig.
1	.286	87.670	4	.000

Canonical Discriminant

At each step the Variable that minimizes the overal Wilk's Lambda is entered.

[1] Maximum number of steps is 52.

[2] Minimum partial F to enter is 3.84.

[3] Maximum partial F to remove is 2.71.

[4] F level, tolerance, or VIN insufficnient for further computation.

Function Coefficients

	Function
	1
Age	.832
Titl	2.098
Uni	.508
Spm	-.001
Constant	-7.187

Unstandardized coefficients

Group covariances of canonical discriminant functions

Deg	Function	1
1.00	1	1.146
2.00	1	.869

Prior Probabilities for Groups

Deg	Prior	Cases Used in Analysis	
		Unweighted	weighted
1.00	.500	35	35.000
2.00	.500	39	39.00
Total	1.000	74	74.00

الفصل السابع

أسلوب دمج طريقتي تحليل الانحدار والمركبات الاساسية في بناء النماذج الاحصائية

Merging Regression & Principal Compnent Analysis for model Building

1-7 الطرق التقليدية والطريقة المقترحة لاختيار افضل طاقم متغيرات

تتوفر عدة طرق ومنهجيات احصائية لغرض اختيار مجموعة المتغيـــرات (التفسيرية) التي يمكن تضمينها في النموذج الذي يتم بناءه، ويأتي هذا الاختيار لاجل تحقيق هدفين حاسمين وهما :

- تقليص عدد المتغيرات التي يضمها النموذج بغية تقليل الكلفة من جهة وتسهيل عملية احتسابه واستخدامه من جهة اخرى .
- الاخر هو تحاشي مشكلة العلاقات الخطية المتداخلة بين المتغيرات التي يتم ترشيحها واخضاعها للتحليل وهي ما يطلق عليها بمسألة "Multicolinearity" .

ومن اهم هذه الطرق هي التي تعتمد الانحدار المتعـــدد Multiple Regression وتحليل المركبات الاساسيةPrincipal Analysis mponent الا ان الطريقة المقترح تطبيقها موضوع هذا الفصل هي حصيلة الدمج بين كلا الطريقتين اعلاه (Daling and Tamura, 1970) والتي يمكن اعتبارها الحالة العملية لطريقة :

Kendall، 1959, Principal Component Regression . والتي من ابرز مزاياها هو توفير مرونة عالية للباحث في اختيار المتغير الاكثر اهمية لتضمينه في النموذج، بالاضافة الى النتائج المعنوية التي يمكن ان تتمخض عنها عملية التحليل بأستخدام المنهجية المقترحة وكما تفصح عنه نتائج الحالة الدراسية لهذا الفصل .

7- 1- 1 استخدام طريقة الانحدار

عند صياغة النموذج الخطي العام، يفترض وجود علاقة خطية بين المتغير المعتمد Y و المتغيـرات المستقلـة اوالتفسيريـة $X_1, X_2, \ldots, X_{k-1}$ والمتغيـر العشوائي ε_i، وان الصيغة العامة لهـذه العلاقـة في حالـة المجتمع هي:

$$Yi = \alpha_0 + B_1 X_1 + B_2 X_2 + \text{-------} + B_{k-1} X_{k}\text{-}1 + \varepsilon_i$$

ان هذا النموذج يجب ان يحقق بالاضافة للفرضيات المتعلقة بالمتغير العشوائي فرضية عدم وجود علاقة خطية ما بين المتغيرات المستقلة (التفسيرية)، اى ان هذه المتغيرات تكون مستقلة عن بعضها كما وان عدد المشاهـدات التـي هي n يجب ان تزيد على عدد المعلمات المجهولة في النموذج اي : $k \leq n$ ، وفي حالة عدم تحقق هذه الافتراضات فأن النتائج التي يتم التوصل اليها لا تكون صحيحة، وبالتالي عدم صلاحية استخدام النموذج الخطي العام الا بعد التأكد اولا من مدى انطباق الافتراضات التي على اساسها تم بناء النموذج على واقع المشاهدات. لانه عندما تكون هناك علاقات تامة او شبه تامة بين المتغيرات المستقلة فان قيمة محدد المصفوفة | $x'x$ | تساوي عندئذ صفر مما يتعذر معه ايجاد $(x'x)^{-1}$ وبالتالي يتعذر تقدير قيم المعالم المجهولة.

7- 1- 2 استخدام طريقة تحليل المركبات الاساسية

وهي الطريقة الاحصائية التي تستخدم مع البيانات ذات المتغيرات المتعددة. وتقوم بتجميع كل مجموعة من البيانات المترابطة خطيا في احد المركبات الاساسية. وتعتمد في ذلك على مصفوفة الارتباط للمتغيرات التي تعتبر الاساس في عملية تحليل المركبات.

وكما تم ذكره في فصل التحليل الوصفي، فان المركبات الاساسية S ' C_p هي متغيرات عشوائية غير مترابطة، وكل منها تضم مجموعة متغيرات عشوائية $(X_1, X_2, \ldots, X_p)$ مترابطة وتشترك بأتجاه خطي، بحيث تأخذ الصيغة التالية :

$$C_j = \sum_{i=1}^{p} a_{ij} \qquad , \qquad j = 1,2,\ldots,p$$

وان a_{ij} هي عبارة عن معاملات عناصر(elements) مصفوفة الموجهات الذاتية الطبيعية (Normalized eigen vectors) لمصفوفة الارتباط للمتغيرات $Xi's$.

وبواسطة مصفوفة قيم الموجهات الذاتية (eigen values) نحصل على الحجم النسبي للتغاير او التباين المفسر للبيانات الاحصائية بواسطة كل من المركبات الاساسية التي يتم احتسابها من المتغيرات $xi's$ بأستخدام التحويل المتعامد (orthogonal transformation) ويطلق على المعاملات aij عادة بتحميلات المركبة (component loadings) وهذه التحميلات تشير الى وزن العلاقة بين المتغيرات $Xi's$ والمركبات الاساسية C_p' S بشرط اخذ الجذر التربيعي لمصفوفة ارتباط تباينات الشيوع لقيم التباين الذاتي (Morrison, 1967). ويطلق على قيم الجذر التربيعي لتباينات الشيوع (Communality) ويمكن الرمز له بـ :

$$h_j = a_j^2 1 + a_j^2 2 + \ldots + a_j^2 p$$

7- 1- 3 استخدام طريقة الدمج بين المركبات الاساسية والانحدار Principal Component Regression Analysis

وتمتاز الطريقة المقترحة بتوفيرها لمرونة كافية لاختيار المتغير الاكثر اهمية من خلال اعطاء الباحث حرية الاختيار وذلك بالاعتماد على خبرته العملية ومثل هذه المرونة لا توفرها الطرق التقليدية التي تعتمد حصرا على الاساليب الميكانيكية بغض النظر عن مدى اهمية المتغير (**Kendall, 1957**).

والفكرة التي تقوم عليها الطريقة هي: عند افتراض ان R هي مصفوفة الارتباط لـ p من المتغيرات التوضيحية او التفسيرية X_i ، وافترضنا بأن C التي هي مصفوفة المركبات الاساسية المتعامدة orthogonal matrix وD هي المصفوفة القطرية Diagonal Matrix لقيم التباين الذاتي λ_i فان :

$$D = \overset{T}{C}'RC$$

المركبات الاساسية هي عناصر للموجه (Vector)

$$= \overset{T}{C}'xg_i$$

وبذلك فإن انحدار المتغير المعتمد Y على المركبات g_i تعطي انحدار تعامدي Orthogonal، ويمكن تلخيص اهم خطوات الدمج بما يلي :

- اختيار عدد المركبات الاساسية التي تفسر اعلى نسبة من التباين ولنقل اكثر من 80٪ مثلا
- تحليل انحدار المتغير التابع (الاستجابة) Y مع كل مركبة يتم اختيارها في الخطوة 1 اعلاه واحتساب قيمة R^2 مع التحقق من صحة اشارة المعاملات.
- اختيار المركبات التي تحقق اعلى R^2.

• اختيار متغير مستقل من كل مركبة تم اختيارها في الخطوة 3 اعلاه على الاسس التالية :

- الاهمية العملية للمتغير بالنسبة للمتغير التابع (الاستجابة)
- سهولة قياسه (كلفته) .
- درجة ارتباطه مع المتغير التابع .

• اجراء عملية تحليل الانحدار للمتغير المعتمد مع كل من المتغيرات المستقلة التي يتم اختيارها في الخطوة الرابعة اعلاه .

2-7 حالة دراسية رقم C_{7-1}

استخدام طريقة الدمج بين المركبات الاساسية والانحدار

▫ سيتم ترشيح ثلاثة عشر متغيرا تم الحصول عليها من مسح احصائي بالعينة شملت 842 مسافرا من مسافري النقل العام بين المدن. والمتغيرات المرشحة هي :

1. مسافة الطريق بالكم X_1 (قيمة مطلقة)
2. يوم السفر X_2 (متغير هيكلي Dummy Variable , ... ,1, 2 ، (7
3. نوع واسطة النقل X_3 (متغير هيكلي : باص متوسط الحجم 1، باص كبير الحجم، 2) ،
4. وقت السفر X_4 (متغير هيكلي : قبل الظهر 1، بعد الظهر 2) ،
5. الغرض من السفرX_5 (متغير هيكلي : رحلة عمل 1، رحلة غير عمل 2)
6. عمر المسافر X_6 (قيمة مطلقة : عدد السنين) ،
7. جنس المسافر X_7 (متغير هيكلي 1،2) ،

8. مهنة المسافر X_8 (متغير هيكلي 1، 2، 3، ...، 5) ،
9. معدل دخل اسرة المسافر الشهري X_9 (قيمة مطلقة : بالدينار) ،
10. طول زمن الرحلة X_{10} (قيمة مطلقة : بالدقائق) ،
11. معدل وقت انتظار المسافر في المحطة X_{11} (قيمة مطلقة : بالدقائق) ،
12. اجور السفر X_{12} (قيمة مطلقة : اجزاء الدينار)
13. معدل عدد رحلات واسطة النقل الواحدة يوميا X_{13} (قيمة مطلقة: عدد)

7-2-1 استخدام منهجية الدمج بين طريقتي الانحدار والمركبات الاساسية

- في الخطوة الاولى وطبقا لمتطلبات تحليل المركبات الاساسية باستخدام برنامج SPSS التي تطرقنا اليها في الفصل الخامس، وكما مبين في الجدول رقم (1.7) حصلنا على 10 مركبات اساسية معنوية استطاعت هذه المركبات ان تقوم بتفسير اكثر من 99٪ من التباين.
- في الخطوة الثانية وعند اجراء عملية تحليل الانحدار باستخدام برنامج SPSS وفقا للخطوات التي تطرقنا اليها في تحليل الانحدار المتعدد في الفصل السادس، بتوظيف كل من المركبات العشرة التي حصلنا عليها في الخطوة الاولى، واحتساب قيمة R^2 لكل منها، نستدل على معنوية 7 مركبات وهي : 10،9،7،6،5،4،1 وكما مبين في الجدول رقم (2.7) .
- **وفي الخطوة الثالثة يتم اختيار متغير واحد من كل مركبة وفقا لمعايير المعنوية ودرجة اهمية المتغير، فيكون لدينا مجموعة المتغيرات التالية :، X4, X11, X12, X9 , X7, X5, X2**

وعند بناء النموذج الذي سيضم مجموعة متغيرات الخطوة الثالثة , سيحقق المعايير الاحصائية التالية :

$$\mathbf{R^2 = 0.84}$$

S.E = 232 F-ratio = 609 Sig at 0.000

عدد المتغيرات = 7

جدول رقم (1.7)

نتائج اختيار طاقم المتغيرات باستخدام طريقة دمج الانحدار وتحليل المركبات الاساسية.

Component	R^2	Variable Selected
1	.05	X_{11}
2	.02	...
3	.00	...
4	.05	X_9
5	.06	X_7
6	.04	X_4
7	.04	X_2
8	.00	...
9	.05	X_5
10	.05	X_{12}

* مع حجم عينة $n = 842$ فأن قيم R^2 في الجدول اعلاه هي عالية المعنويـــة أي $\alpha = 0.01$

7-2-2 نموذج الانحدار بطريقة شمول كافة المتغيـرات

All Possible Regression

بأفتراض ان البيانات خالية من مسألة العلاقات المتداخلة (Multicollinearity) وتضمين جميع المتغيرات المستقلة البالغ عددها 13 متغيرا في النموذج باستخدام طريقة enter، فأن النموذج سيؤول الى النتائج التالية:

$$R^2 = 0.84$$

S.E. = 231 F = 329 Sig. at 0.000

عدد المتغيرات = 13

7-2-3 نموذج تحليل المركبات الاساسية

وبأستخدام طريقة تحليل المركبات الاساسية، نستدل وكما موضح في الجدول (2.7) بأن هناك 10 مركبات (عوامل) معنوية، استطاعت تفسير 99.4٪ من مجموع التباين. وبأعتماد حجم التحميل (size of loading) واهمية المتغيرات بالنسبة للمتغير المعتمد فأن هذه المركبات العشرة يمكن تسميتها كما يلي :

1. **طول زمن الرحلة**: ان اعلى معامل تحميل في هذه المركبة يعود لمتغيري طول مسافة الطريق (0.98) وطول زمن الرحلة (0.97)، وقد تم اختيار متغير طول زمن الرحلة لتمثيل المركبة كون المتغير يتصف بمرونة اكبر عند توظيفه في التخطيط كونه يدخل ضمن متغيرات النظامPolicy variables التي تتيح لمتخذ القرار استخدامة في عملية التطوير. وان نسبة التباين التي قامت مجموعة متغيرات هذه المركبة بتفسيرها هي 33.9٪. وبنفس الطريقة يتم التامل في متغيرات كل مركبة في تسمية المتغير الذي يتم اختاره كممثل للمركبة، فتم اختيار المتغيرات التالية :
2. **واسطة النقل**: وقد بلغ معامل التحميل لهذا المتغير هو 99٪ وساهمت بتفسير 13.7٪من مجموع التباين .
3. **عمر المسافر**: بمعامل تحميل مقداره 97٪، وتفسير 8.8٪ من مجموع التباين.
4. **معدل دخل الاسرة الشهري**: بمعامل تحميل مقداره 98٪ والمساهمة بتفسير 7.9٪ من مجموع التباين .
5. **جنس المسافر**: بمعامل التحميل مقداره 99٪ وبتفسير 7.4٪ من مجموع التباين .

6. **الغرض من السفر**: بمعامل تحميل مقداره 99٪ وبتفسير 6.6٪ من مجموع التباين .

7. **يوم السفر**: بمعامل تحميل مقداره 99٪ وبتفسير 6.5٪ من مجموع التباين.

8. **مهنة المسافر**: بمعامل تحميل مقداره 95٪ وبتفسير 5.6٪ مـن مجموع التباين

9. **وقت السفر**: بمعامل تحميل مقداره 95٪ وبتفسير 5.1٪من مجموع التباين.

10. **معدل وقت الانتظار في المحطة**: بمعامل تحميل مقداره 93٪ وبتفسير 3.9٪ من مجموع التباين.

جدول رقم (2.7)

معاملات تحميل تحليل المركبات الاساسية للمتغيرات المستقلة الوضعية

Variable	**Components**									
	C_1	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**	C_{10}
X1	.98									
X2							.99			
X3		.99								
X4	-.31									
X5						.99			.95	
X6			.97							
X7					.99					
X8								.95		
X9				.98						
X10	.97									
X11	.35									
X12	.94									
X13	.95									
Eigen value	4.4	1.76	1.15	1.03	.97	.86	.84	.72	.66	.51
% Variation	33.9	13.7	8.8	7.9	7.4	6.6	6.5	5.6	5.1	3.9
Cum%ofVar.	**33.9**	**47.9**	**56.4**	**64.3**	**71.7**	**87.3**	**84.8**	**90.4**	**95.5**	**99.4**

والمعايير التالية توضح تقييم نتائج عملية بناء النموذج الذي يتضمن مجموعة لـ 10 متغيرات التي تم اختيارها من المركبات اعلاه .

$$R^2 = 0.84$$

$$S.E = 231$$

$$F = 609 \qquad \text{Sig. at } 0.000$$

عدد المتغيرات = 10

7- 2- 4 المقارنة بين النتائج

لاجل اعطاء صورة واضحة من خلال المقارنة فأن الجدول رقم (3.7) التالي يعطي حصيلة كل طريقة ومستوى معنوية كل نموذج .

جدول رقم (3.7)

يوضح المقارنة بين نتائج الحالات الثلاث للتحليل

المعايير الاحصائية	النموذج النهائي لكافة المتغيرات بطريقة الانحدار	النموذج النهائي بالاعتماد على طريقة تحليل المركبات	النموذج النهائي بالاعتماد على طريقة الدمج بين تحليل المركبات والانحدار
No. of Variables	13	10	7
R^2	0.84	0.84	0.84
S.E	231	233	232
F-ratio	329	609	856
Sig. At	0.000	0.000	0.000

ومن الجدول نستدل بأنه رغم ما حققته المنهجية المقترحة وهي طريقة الدمج بين تحليل المركبات والانحدار من انخفاض في عدد المتغيرات التي يتضمنها النموذج الذي يعتبر ميزة اقتصادية مهمة، فانها استطاعت ايضا تحقيق مايلي:

- زيادة في درجة المعنوية بالنسبة لمعيار F-ratio مع عدم حصول اى انخفاض في درجة معنوية باقي المعايير الاحصائية، مما يدل على كفائتها ومعنوية نتائجها.
- تحديد اهم المتغيرات وفقا للاسباب المنطقية التي لها علاقة في تفسير الطلب مع توفير حرية للباحث في اختيارالمتغيرات المهمة. بالاضافة الى توفير صورة عن المتغيرات التي لم يتضمنها النموذج ولكنها تصبح ممثلةProxy بواسطة المتغيرات الداخلة في النموذج استنادا الى العلاقة الخطية التي تربط هذه المتغيرات في كل مركبة .
- ان المنهجية تزداد فائدتها ومردوداتها الاقتصادية عندما نكون امام عدد كبير من المتغيرات وذلك لما تيسره في الكشف عن تلك الاكثر اهمية من خلال التأمل بحجم معاملات التحميل لكل متغير ودرجة اهميته للضاهرة تحت الدراسة .
- وفقا للاسس النظرية لتحليل المركبات الاساسية القائلة بوجود علاقات خطية مترابطة بين متغيرات كل مركبة وعدم وجود العلاقة بين المركبات. فأن عملية التخلص من العلاقات المتداخلة بين المتغيرات هي اكثر وضوحا وفعالية وتيسرا من خلال اختيار الممثل الانسب في كل مركبة .

الفصل الثامن

اختبار الفروض وتحليل التباين

Hypothsis Testing and Analysis of Vriance

8- 1 مفهوم وخصائص الاختبار

Definition and Properities

وهو من المواضيع الواسعة الاستخدام ويعتبراحد المواضيع الرئيسية للاستدلال الاحصائي ويدخل بصورة خاصة تحت موضوعي التوزيعات الاحتمالية Probability Distributions و توزيع المعاينة Sampling Distribution، واستخدامه يستهدف الوصول الى قرار بشان قبول او رفض فرضية محددة وفقا لمعطيات العينة المتوفرة لدى متخذ القرار. ويمكن اجمال اهم اهداف عملية الاختبار بما يلي :

- تقدير معلمة المجتمع بالاعتماد على معطيات العينة المسحوبة منه للتوصل الى درجة الاعتمادية على نتائج العينة في تمثيلها للمجتمع، ولتقريب الصورة نفترض بان شركة ما تريد التاكد من ان انتاجها مطابق للمواصفات المقررة، فتقوم بسحب عينة واختبار نتائجها لمعرفة ان كانت فعلا قد مثلت هذه العينة المواصفات المقررة في العملية الانتاجية ، او ان تدعي الشركة بان هكذا مواصفات موجودة في انتاجها، وتقوم جهة بحثية او حكومية مختصة بسحب عينة لاختبار صحة ادعاء الشركة من عدمه وهكذا .
- اختبار الفروق بين النتائج الفعلية للعينة والنتائج الفرضية المتوقعة، وهذه الفروق قد تكون نتيجة فروق زمنية او مكانية او ظروف معينة سواء

اكان هذا يتعلق بسلع او خدمات اوغيرها من الانشطة المتماثلة ومثل هذه الفروق قد تظهر ايضا في نفس الزمن و ذات المكان على نطاق فروع تعود لنفس البنك او المنظمة او المؤسسة التي تمارس نشاط مالي او اجتماعي او انتاجي او غيره.

واهم الاسس التي تقوم عليها عملية الاختبار هذه، هي :

8- 1- 1 الفروض Hypotheses

وتتمثل بفرضيتين الاولى تدعى فرضية العدم Null hypotheses ويرمز لها عادة H_0 وهي تتضمن الهدف المطلوب اختباره، ففي حالة قبولها يعني انها متوافقة مع الهدف، اي عدم وجود ما يدعو الى رفض النتائج. والثانية وتسمى بالفرضية البديلة Alternative hypotheses ويرمز لها H_1 ، فعند رفض H_0 يعني قبول H_1 والعكس صحيح. وتاخذ الفرضيات الشكل التالي :

$$H_0 : \mu = \overline{X}$$

$$H_1 : \mu \neq \overline{X}$$

فمثلا اذا كنا بصدد اختبار من ان معدل وزن الطالب في المجتمع هو 65 كغم او اقل، فان الفرضية ستكون كالاتي :

$$H_0 : \mu \leq 65$$

$$H_1 : \mu > 65$$

او اختبار من نسبة خاصية معينة في المجتمع تساوي 0.02، فان شكل الفرضية سيكون :

$$H_0 : P = 0.02$$

$$H_1 : P \neq 0.02$$

8- 1- 2 الاخطاء Errors

(1) الخطأ من النوع الاول Type I error

فعند رفض فرضية العدم H_0 ولكن كان يجب قبولها لان عملية الرفض هو نتيجة خطأ في المعطيات، عندها نقع في الخطأ من النوع الاول وان احتمال الوقوع في مثل هذا الخطأ هو α وتدعى بمستوى الدلالة (المعنوية) Level of Significant وكما موضح في الفقرة (8-1-3)، وكلما تقل قيمة α يقل احتمال الوقوع في الخطأ من النوع الاول .

(2) الخطأ من النوع الثاني Type II error

ويقع في حالة قبولنا لفرضية العدم H_0 بينما كان يجب رفضها، وان احتمال الوقوع في هذا الخطأ يرمز له β ويدعى بقوة الاختبار Testing Power وكما موضح في الفقرة (8-1-4).

ولتقريب صورة وقوع هذه الاخطاء : لنفترض بان متوسط استهلاك الاسرة من القهوة في مدينة ما وفقا لمعطيات عينة هو $\bar{x} = 150$ غم شهريا، وعلى فرض بان المتوسط الحقيقي لاستهلاك القهوة في المجتمع هو $\mu = 146$ غم شهريا. فان الاختبار سيعستمد على مقدار الفرق بين متوسطي العينة والمجتمع والذي هو ± 4 ، وعليه فمن المحتمل الوقوع باحدى الحالتين التالية:

الحالة الاولى : هو ان العينة قد تضمنت نسبة اعلى من الاشخاص من ذوي الاستهلاك العالي للقهوة، وبالتالي جاء متوسطها اعلى من الواقع، عندها سنقع في الخطأ من النوع الاول α، وهو استنتاج خاطئ .

الحالة الثانية : قد يكون الفرق بين متوسطي العينة والمجتمع هو صحيح نتيجة شمول العينة على نسبة اعلى من الاشخاص من ذوي الاستهلاك المنخفض وبالتالي جاء متوسطها صغير ومقارب لمتوسط المجتمع، لكن بسبب اخطاء المعاينة ظهر لنا بان الفرق صغير و غير معنوي، عندها يكون الاستنتاج خاطئ فنقع في الخطأ من النوع الثاني β .

ان تقليص احتمال الخطأ من النوع الاول يمكن ان يتم من خلال رفع قيمة مستوى المعنوية α فنجعلها مثلا 0.05 بدلا 0.01، الان ذلك يرفع من احتمال الوقوع في الخطأ من النوع الثاني، والعكس صحيح فان تقليص احتمال الوقوع في الخطا من النوع الثاني يزيد من احتمال الوقوع في الخطا من النوع الاول. لذا فالامر يتطلب مراعاة اي من الخطئين يشكل خطورة اعلى. فالوقوع في الخطا الاول يعني جصول زيادة في ضخ مادة القهوة الى السوق، في حين ان الوقوع في الخطا من النوع الثاني سيؤدي الى شحة في عرض القهوة .

8- 1- 3 مستوى المعنوية α Level of Significance

وهي تمثل الحد الاعلى لاحتمال الوقوع في الخطأ من النوع الاول Type I error، وبذلك فان مستوى المعنوية α هي تعين منطقة (مساحة الرفض تحت منحنى توزيع اختبار الاحصاءة مثل t او f ...الخ. وعادة ما تستخدم القيم 0.01 و 0.05 و0.10 كمستوى معنوية .

8- 1- 4 قوة الاختبار β Testing Power

ان حالة قبولنا لفرضية العدم H_0 وهي غير صحيحة سيؤدي للوقوع في الخطأ من النوع التاني Type II error، ، ويعتمد احتمال الوقوع في هذا الخطأ على مقدار الابتعاد (مستوى المعنوية α) عن H_0 ، وعلى حجم العينة n، وعلى الانحراف المعياري للمجتمع σ، ونوع الاختبار ان كان من

جانب واحد او من جانبين (موضوع الفقرة 8-1-5 التالية)، وان صيغة حسابه هي :

$$\beta = \sqrt{\frac{n(\mu - \mu_o)}{\sigma}}$$

فمثلا لو كان لدينا : $n=5000$, $\sigma=11$, $\mu=25.6$, $\mu_o=25$ ، فان قوة الاختبار β او احتمال رفض فرضية :

$$H_o : \mu_0 = 31$$

يتم بعد ايجاد قيمة متوسط العينة $\bar{x}$ الذي يؤدي الى رفض H_o وكالاتي :

فالخطأ المعياري ل $\bar{x}$ هو :

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{11}{\sqrt{5000}} = 0.155$$

وعند مستوى دلالة $\alpha = 0.05$ نرفض H_o اذا وقع $\bar{x}$ خارج :

$$25 \pm (1.96)(0.155)$$
$$\pm 0.30425$$

اي ان قرار رفض H_o هو اما :

$$\bar{x} < 24.696$$

او

$$> 25.304\bar{x}$$

وان احتمال وقوع $\bar{x} < 24.696$ هو :

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{24.696 - 25.6}{0.155} = -5.832$$

ومن الملحق (6) نجد : $p\ (0 \text{ to } -5.83) = 0$

واحتمال وقوع $\bar{x} > 25.304$ هو :

$$z = \frac{25.304 - 25.6}{0.155} = -1.91$$

ومن الملحق رقم (7) نجد : $p\ (0 \text{ to } -1.91) = 0.9719$

اذن ان قوة الاختبار β هي $0.9719 = 0.09719 + 0$ بكلمة اخرى ان وقوع الخطأ من النوع الثاني يكون عند احتمال $= 1 - 0.9719 = 0.0281$

8- 1- 5 اختبار من جانب واحد ومن جانبين

I tail test & II tails test

ويقصد به اتجاه الانحراف عن فرضية العدم هو باتجاه واحد او باتجاهين موزع على جانبين، وهو ما يعتمد على صيغة فرضية العدم، فاذا كانت الاشارة هي ≥ (اكبر من او يساوي) او ≤ (اقل من او يساوي) أي :

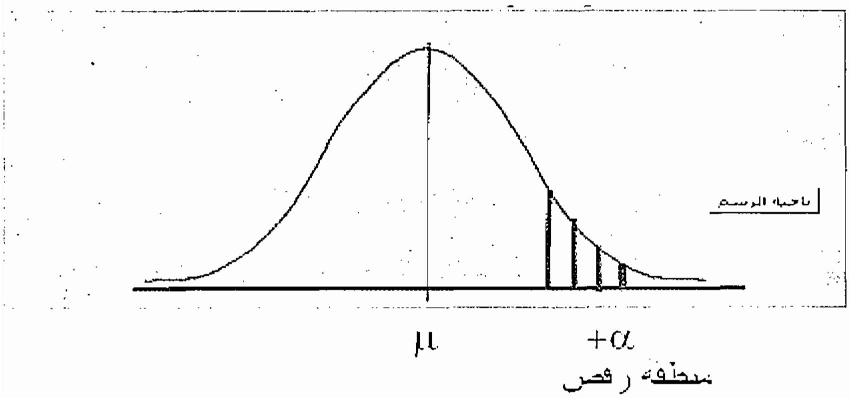
$$H_o : \mu \geq \bar{x}$$

$$H_1 : \mu < \bar{x}$$

فهذا يعني بان الاختبار من جانب واحد لانه في حالة رفض الفرضية فمن المتوقع حصرا بان الفرضية البديلة سيكون معلوما اتجاهها، فاذا كان الاتجاه موجب مثلا فسيكون كما مبين في الشكل البياني رقم (1.8) .

شكل بياني رقم (1.8)

يوضح وقوع الخطأ α على جانب واحد



اما في الحالة التي تكون فيها فرضية العدم H_0 مع اشارة يساوي =، فان التوقع في حالة رفضها هو اما ستكون :

$$H_1: \mu > \bar{x}$$

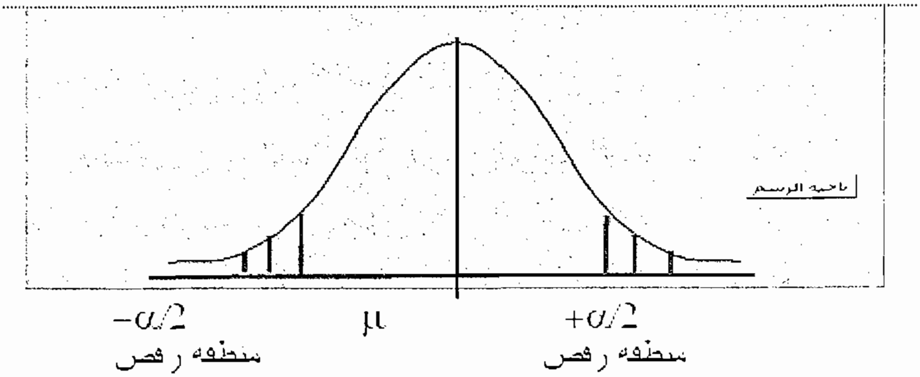
أو

$$H_1 : \mu < \bar{x}$$

اي عدم معلومية الاتجاه الذي ستكون عليه نتيجة الاختبار مسبقا، وبذلك سيتوزع الانحراف (الخطا) على جانبي التوزيع، وكما هو مبين في الشكل البياني رقم (2.8) .

شكل بياني رقم (2.8)

يوضح توزيع الخطأ α على جانبين α/2



8- 1- 6 اتخاذ القرار بشان نتيجة الاختبار

Decision Making

ان قرار قبول او رفض فرضية العدم H_0 يعتمد على نتيجة مقارنة القيمة المحتسبة مع القيمة الجدولية تحت مستوى المعنوية α المقرر، فاذا كانت القيمة المحتسبة تقع في منطقة الرفض، اي انها اقل من القيمة الجدولية

عندها نقبل فرضية العدم ويصبح استنتاجنا مطابق لمنطوق الفرضية. في حين نرفض فرضية العدم H_0 اذا كانت القيمة المحتسبة اكبر من القيمة الجدولية تحت مستوى المعنوية المقرر واللجوء الى قبول الفرضية البديلة H_1.

8- 2 اختبار المتوسطات Testing of Means

8-2-1 الاختبار الاحادي (متوسط مجتمع واحد) One Sample test

ويقصد به اختبار متوسط العينة $\bar{x}$ (او القيمة X) مع متوسط المجتمع μ للتوصل ان كان هناك فرق جوهري بينهما وعلى افتراض تساوي التباين لكلاهما. مثال ذلك اختبار اداء احد فروع بنك ما مع اداء البنك الرئيسي الذي يعود اليه، او اختبار متوسط عينة من منتجات شركة ما للتاكد من مطابقة الانتاج لمواصفات انتاج الشركة المقررة وهكذا.

(1) خصائص وأجراءات الاختبار الاحادي

ومنطوق فرضية العدم H_0 هو ان متوسط المجتمع μ مساويا لمتوسط العينة $\bar{x}$ (او لمتوسط فرضي μ_0)، وعلى اعتبار ان المتغير العشوائي $\bar{x}$ عبارة عن متوسط متوسطات العينات، وان الانحراف المعياري للمتغير هو لمتوسط العينات $\sigma_{\bar{x}}$ ايضا، اي :

$\mu_{\bar{x}} = \mu$ وان : $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ في حالة معلومية تباين المجتمع σ^2

او $\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$ في حالة مجهولية تباين المجتمع σ^2

حيث ان σ هو الانحراف المعياري للمجتمع المسحوبة منه العينة .

ومن خلال تحويل قيم المتغير العشوائي x_i الى قيم طبيعية معيارية، Z او t لنحصل على المنطقة الحرجة، باستخدام الصيغة التالية :

$$Z = \frac{\mu_{\bar{x}} - \mu}{\frac{\sigma}{\sqrt{n}}}$$ او عند المقارنة مع متوسط فرضي : $$Z = \frac{\mu - \mu_0}{\sigma}$$

او في حالة مجهولية تباين المجتمع، وهي الحالة الغالبة الاستخدام عمليا

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}.$$

حيث ان s هو الانحراف المعياري للعينة .

وبالرجوع الى الجدول في الملحق (2) نجد القيمة الجدولية لـ Z والملحق (3) لايجاد قيمة t الجدولية عند مستوى معنوية α محدد، وان كان الاختبار من جانب واحد او من جانين وفقا لطبيعة الفروض، لاجل اتخاذ قرار الرفض او القبول .

مع التنويه الى انه في حالة عدم معلومية توزيع المجتمع وكان حجم العينة $n \geq 30$ فنفترض دائما بانها مقاربة للتوزيع الطبيعي، وعليه فان قيم كل من z و t تكون متقاربة .

مثال (1.8) : مصنع لانتاج معدات رياضية ادعى بانه استطاع صناعة مضرب للتنس بمقاومة متوسطها $\mu = 6.5$ كغم، والمطلوب اختبار صحة ادعاء المصنع مع نتائج عينة حجمها $n = 62$ تم سحبها من انتاج المصنع والمبينة قيمها في ادناه عند مستوى معنوية $\alpha = 0.01$.

6.7, 6.7, 6.6, 6.4, 5.9, 6.5, 7.1, 7.0, 6.5, 6.5, 6.0, 6.3, 6.4, 6.0, 6.7, 5.9, 5.8, 6.8, 6.4, 5.9, 5.8, 6.8, 6.4, 5.9, 7.1, 7.0, 6.5, 6.5, 6.0, 6.3, 6.4, 6.5, 6.7, 5.9, 6.7, 7.1, 7.0, 5.8, 6.7, 6.3, 6.7, 6.3, 6.1, 6.9, 6.8, 5.9, 6.7, 6.5, 6.4, 5.6, 7.2, 7.0, 6.8, 6.6, 6.6, 6.1, 6.5, 5.9, 6.7, 6.4 , 6.3, 6.4 .

الحل لـ **(1.8)** : لدينا : $\mu = 6.5$، $s = 0.54$، $\bar{x} = 6.471$، $n = 62$

- تحديد الفرضية المستهدفة :

$$H_0 : \mu = 6.5$$
$$H_1 : \mu \neq 6.5$$

- حيث ان اشارة الفرضية البديلة هي عدم المساواة $\neq$ فان الاختباريكون من جانبين، اي :

$\alpha/2 = 0.005$، ومن الملحق (5.3) نجد ان القيمة الجدولية لـ $t_{\alpha/2} = 2.66$

- بتطبيق صيغة حالة مجهولية تباين المجتمع لـ t ، نحصل على :

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{6.471 - 6.5}{\frac{0.54}{\sqrt{62}}} = 0.42274$$

- القرار : حيث ان قيمة t المحتسبة البالغة 0.42274 هي اقل من القيمة الجدولية $t_{\alpha/2} = 2.66$، عليه نقبل فرضية العدم ونستدل على صحة ادعاء المصنع .

(2) حالة دراسية رقم $C_{8\text{-}1}$

استخدام برنامج SPSS لانجاز الاختبار الاحادي بتوظيف معطيات المثال (1.8)

عقب انشاء ملف بالمعطيات اعلاه، يتم متابعة الخطوات التالية :

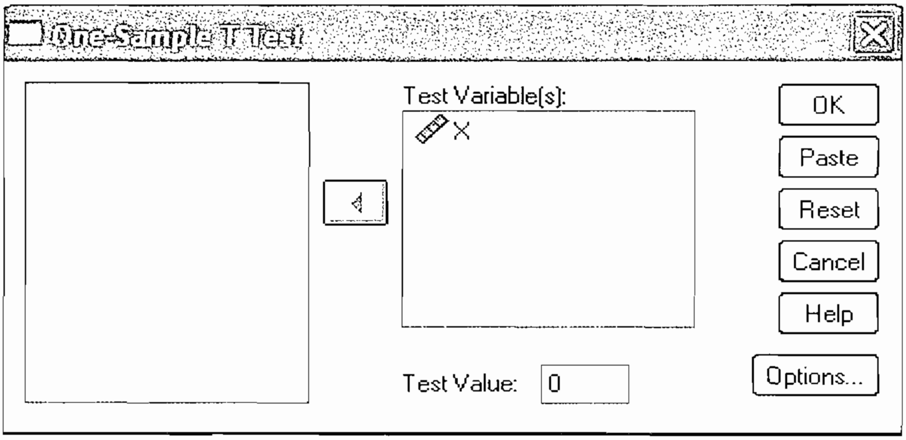
- استخدام قائمة Analysis من برنامج SPSS ومن ثم الامر الفرعي Compare Means ومنه اختيار One sample test ،
- يظهر مربع الحوار One sample test المبين في الشكل البياني رقم (3.8)، ويتم نقل المتغير الى داخل المربع تحت Test Variable باستخدام السهم الجانبي ،

- الكبس على ايقونة Options من اجل تحديد درجة الثقة المستهدفة، لتظهر لنا اللوحة المبينة في الشكل البياني رقم (4.8)، وبعد الانتهاء من تدوين درجة الثقة في حالة كانت تختلف عن القيمة المثبتة وهي 0.95 يتم الكبس على ايقونة Continue للعودة الى مربع الحوار،
- الكبس على ايقونة Ok الموجودة على مربع الحوار لنحصل على المخرجات المبينة في مجموعة الجداول رقم (1.8) .

ومن المخرجات المبينة في الجدول رقم (1.8) نستدل على صحة ادعاء الشركة، حيث ان النتائج مقبولة بمعنوية عالية $\alpha = 0.000$ وان متوسط المجتمع μ عند درجة ثقة مقدارها 99 ٪ يقع بين القيمتين 6.653 و 6.2889 وان متوسط العينة 6.471 شبه مطابق للمتوسط الذي اشارت اليه الشركة والبالغ 6.5 كغم. وبذلك تقبل فرضية العدم H_0 وهي : $H_0:$ $\mu = 6.5$ ورفض الفرضية البديلة : $H_1: \mu \neq 6.5$. مع الاشارة الى ان قيمة المحتسبة المبينة في الجدول اعلاه تعني 0.945 .

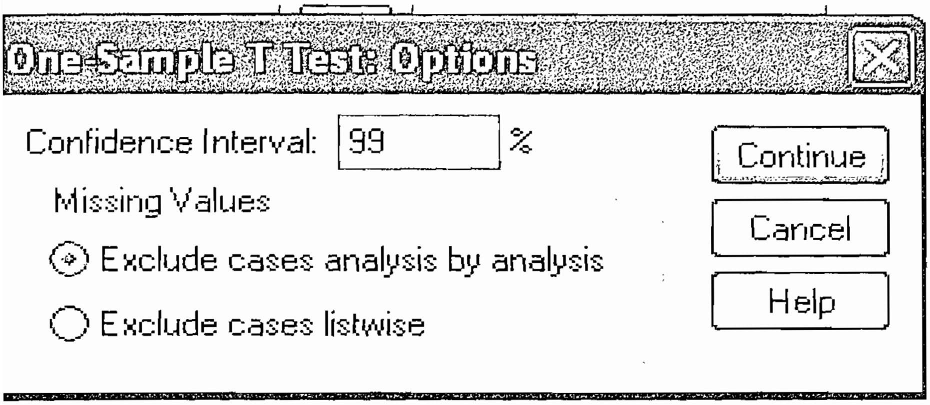
الشكل البياني رقم (3.8)

مربع حوار اختبار One Sample T-test



الشكل البياني رقم (4.8)

لوحة Options لتدوين درجة الثقة المستهدفة One Sample T-test



مدموعة جداول رقم (1.8)

مخرجات نتائج تحليل الاختبار الاحادي One sample test

	N	Mean	Std. Deviation	Std. Error Mean
X	62	6.4710	.53909	.06846

One-Sample Test

	t	df	Sig. (2-tailed)	Mean Difference	99% Confidence Interval of the Difference	
					Lower	Upper
X	94.516	61	.000	6.47097	6.2889	6.6530

(3) اختبار نسبة خاصية معينة P

وهي الحالة يكون المطلوب فيها اختبار نسبة p بدلا من اختبار متوسط ، ويصل ذلك مع الظواهر التي يتم قياسها من خلال تقدير نسبة وقوعها، كما في نسبة الاميين او نسبة الحاصلين على شهادة بمستوى معين او نسبة وحدات الانتاج الصالحة ..... الخ. وكما لاحظنا في موضع التوزيع الاحتمالي الثنائي، يرمز لنسبة وقوع الظاهرة في المجتمع بـ P وعدم وقوعها بـ Q والتي هي عبارة عن Q = 1-P، وبذلك ففي حالة معلومية تباين المجتمع تصبح صيغة الاختبار الاحادي One Sample test كالاتي :

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

حيث ان PQ هو تباين المجتمع الذي تعود اليه النسبة .

مثال (2.8) : يتوفر في الاسواق دواء، على اساس ان نسبة نجاحه في تخفيض توتر الاعصاب هي 0.6، وظهر دواء جديد لنفس المرض كان قد تم تجربته على عينة تتكون من 100 شخص، ودلت النتائج على شفاء 70 شخص منهم عند استخدام هذا الدواء الجديد. فهل يمكن الاستدلال على ان الدواء الجديد هوافضل من النوع المتوفر في الاسواق عند مستوى معنوية 0.05 .

الحل لـ (2.8) :

نحدد الفرضية :

$$H_0 : P > 0.6$$

$$H_1 : P < 0.6$$

▪ استخدام الملحق رقم (4.3) لايجاد قيمة Z الجدولية عند مستوى معنوية 0.05، وحيث ان الاختبار من جانب واحد كما يتضح من الفرضية، فان : $Z_{0.05} = 1.64$

▪ وباستخدام الصيغة اعلاه نحصل على :

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$$Z = \frac{0.7 - 0.6}{\sqrt{\frac{(0.6)(0.4)}{100}}} = 2.04$$

▪ وحيث ان قيمة Z المحتسبة هي اكبر من القيمة الجدولية $Z_{0.05} = 1.64$، عليه نرفض H_0 ونستدل على ان نسبة الدواء الجديد ليس افضل من الدواء المتوفر في الاسواق .

8- 2- 2 اختبار الفرق بين مجتمعين مستقلين (متوسطي عينتين مستقلتين) Independent samples T-test

(1) خصائص واجراءات اختبار الفروق بين مجتمعين مستقلين

ويهف الاختبار معرفة ان كان الفرق بين متوسطي العينتين المسحوبتين من مجتمعين مستقلين يعود الى الصدفة او ان الفرق جوهري، كاختبار مستوى جودة عينتين من منتجات صنف ما لشركتين مستقلتين عن بعضهما، او لظاهرة محددة لبلدين مختلفين وهكذا .

❖ **حالة معلومية تبايني المجتمعين الموزعة طبيعيا ،**

يعتمد الاختبار على ان توزيع العينتين للفرق $\bar{x}_1 - \bar{x}_2$ هو مقارب للتوزيع الطبيعي للمجتمعات المسحوبة منها والتي الفرق بين متوسطيها هو $\mu_1 - \mu_2 = 0$ ، وانحرافها المعياري :

$$\sigma_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1}{\sqrt{n_1}} + \frac{\sigma_2}{\sqrt{n_2}}$$

وان صيغة الاختبار التي تستخدم هي :

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}}$$

حيث ان :

$(\mu_1 - \mu_2) = 0$ من الناحية النظرية،

❖ **حالة مجهولية قيم σ_1 و σ_2 للمجتمعات الموزعة طبيعيا المسحوبة منها العينتين، وهنا نواجه حالتين هما :**

- اما ان يكون تبايني المجتمعين الموزعين طبيعيا المجهولين متساويين، عندها نستخدم الصيغة التالية :

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

مع درجات حرية عددها n_1+n_2-2 ،وحيث ان الفرض هو ان التباينين متساويين فان :

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

- او في حالة مجهولية وعدم تساوي تباين المجتمعين الموزعين طبيعيا، فتكون صيغة الاختبارهي:

$$t' = \frac{(\bar{x}_1 - \bar{x}_2) - (u_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

وفيها تكون القيمة الجدولية عندما يكون الاختبار او المنطقة الحرجة من جانبين مقاربة الى :

$$t'_{1-\alpha/2} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$$

حيث ان : $w_1 = \frac{s_1^2}{n_1}$ ، $w_2 = \frac{s_2^2}{n_2}$ ، وان :

$t_1 = t_{1-\alpha/2}$ مع درجات حرية n_1-1

$t_2 = t_{1-\alpha/2}$ مع درجات حرية n_2-1

و القيمة الجدولية في حالة الاختباراو المنطقة الحرجة من جانب واحد، فتكون مقاربة الى :

$$t'_{1-\alpha} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$$

حيث ان : w_1 , w_2 كما في اعلاه، وان :

$t_1 = t_{1-\alpha}$ مع درجات حرية n_1-1

$t_2 = t_{1-\alpha}$ مع درجات حرية n_2-1

مثال (2.8) : المطلوب اختبار الفرضية القائلة من الن مجتمعين موزعين طبيعيا يختلفان في قيمة وسطهما لدخل الفرد الشهري عند مستوى معنوية $\alpha = 0.05$، وان حجم العينة المسحوبة من المجتمع الاول هو $n_1=10$ ، وحجمم العينة الثاني هو $n_2=20$، وان قيم وسطيهما وانحرفهما المعياري هو : $\bar{x}_1 = 62.6, s_1 = 33.8$ ، $\bar{x}_2 = 47.2, s_2 = 10.1$

الحل ل (2.8) :

- نحدد الفرضية :

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 - \mu_2 = 0$$

- كون ان قبول H_1 يتحقق في حالتي اكبر واقل، فان الاختبار يكون من جانبين، وباستخدام الصيغة اعلاه لايجاد القيمة الجدولية نحصل على :

$$w_1 = \frac{s_1^2}{n_1} = \frac{(33.8)^2}{10} = 114.244$$

$$w_2 = \frac{s_2^2}{n_2} = \frac{(10.1)^2}{20} = 5.1005$$

- ومن الملحق (5.3) نجد ان :

$$t_1 = t_{1-\alpha/2} = 2.262$$

$$t_2 = t_{1-\alpha/2} = 2.093$$

- وبالتعوض في الصيغة التالية يكون لدينا :

$$t'_{1-\alpha/2} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2} = \frac{114.244(2.262) + 5.1005(2.093)}{114.244 + 5.1005} = 2.255$$

$$t' = \frac{(\bar{x}_1 - \bar{x}_2) - (u_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(62.6 - 47.2) - 0}{\sqrt{\frac{(33.8)^2}{10} + \frac{(10.1)^2}{20}}} = 1.4$$

- وبتطبيق صيغة عدم المساواة في التباين نحصل على :

- القرار : وحيث ان قيمة t' المحتسبة هي اقل من قيمة $t'_{1-\alpha/2}$ الجدولية، عليه نقبل H_0 ونستدل على عدم وجود فرق جوهري بين معدل دخل الفرد الشهري للمجتمعين .

مثال (3.8) : المطلوب اختبار عند مستوى معنوية $\alpha = 0.05$، ان كان هناك فرق في عمر الطفل عند المشي لمجتمعين، جمعت عينة من كل منهما بالاشهر وكما مبين في الاتي :

n_1= 9.0, 10.1, 9.2, 10.2, 10.0, 12.8, 13.4, 8.7, 10.5, 11.1

n_2= 9.5, 12.3, 13.2, 12.6, 13.4, 9.6, 9.8, 12.2, 12.0, 10.2

الحل لـ (3.8) :

<u>اولا :</u> على فرض تساوي تباين المجتمعين :

لدينا :

$$\sum x_1 = 106.3, n_1 = 10, \bar{x}_1 = 10.63, s_1 = 1.35$$

$$\sum x_2 = 115.8, n_2 = 10, \bar{x}_2 = 11.58, s \ = 1.54$$

- نحدد الفرضية :

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 = 0$$

- حيث ان قبول H_1 يكون عند حالتي اكبر او اقل، فيكون الاختبار من جانبين، وباستخدام الملحق رقم (5.3) عند مستوى معنوية $\alpha/2 = 0.05/2$ مع درجات حرية عددها 18، فان القيمة الجدولية هي :

$$t_{0.025} = 2.101$$

- وبتطبيق الصيغة التالية نحصل على :

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{9(1.823) + 9(2.362)}{18} = 2.092$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

$$= \frac{10.62 - 11.56 - 0}{\sqrt{\frac{2..092}{10} + \frac{2.092}{10}}} = -1.469$$

▪ القرار : حيث ان قيمة t المحتسبة هي اقل من القيمة الجدولية $t_{0.025} = 2.101$، عليه نقبل H_0 ونستدل من انه ليس هناط فرق جوهري بين معدل عمر الطفل عند المشي لكلا المجتمعين .

ثانيا : على فرض عدم تساوي التباين غير المعلوم :

عندها يتم استخدام صيغة هذه الحالة وهي :

$$t' = \frac{(\bar{x}_1 - \bar{x}_2) - (u_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

وبالتعويض بالقيم اعلاه نحصل على :

$$t' = \frac{(10.62 - 11.56) - 0}{\sqrt{\frac{1.823}{10} + \frac{2.362}{10}}}$$

$$= \frac{-0.94}{0.647} = 1.4528$$

وهي مقاربة جدا للنتيجة التي تم الحصول عليها لحالة فرضية تساوي تباين المجتمعين .

مثال (4.8) : المطلوب اختبارعند مستوى معنوية $\alpha = 0.05$، من ان متوسط العينة الاولى التي حجمها $n_1 = 35$ هو $\bar{x}_1 = 360$ وانحرافها المعياري $s_1 = 25$ هو اكبر من من متوسط العينة الثانية الذي هو $\bar{x}_2 = 350$ وحجمها $n_2 = 36$ وانحرافها المعياري $s_2 = 30$.

الحل لـ (4.8) :

- نحدد الفرضية :

$$H_0 : \mu_1 - \mu_2 \geq 0$$
$$H_1 : \mu_1 - \mu_2 \lhd 0$$

- وحيث ان الاختبار هو من جانب واحد كما نستدل من منطوق الفرضبة، وان القيمة الجدولية هي : $t'_{0.05},69 = 1.667$

- باستخدام صيغة التباينين مجهولين نحصل على :

$$t' = \frac{(\bar{x}_1 - \bar{x}_2) - (u_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(360 - 350) - 0}{\sqrt{\frac{625}{35} + \frac{900}{36}}} = \frac{10}{9.396} = 1.064$$

- القرار : حيث ان القيمة المحتسبة هي اقل من القيمة الجدولية $t'_{0.05},69 = 1.667$، عليه نقبل فرضية العدم، ونستدل على ان متوسط العينة الاولى هو اكبر من متوسط العينة التانية .

مثال (5.8) :

قام احد الباحثين بجمع عينتين لاحد انواع منتجات المواد الغذائية المعلبة من مصنعين في بلدين مختلفين، وذلك بهدف اختبارتحقيق الوزن المقرر البالغ $\mu = 50$ غم، وكان حجم العينة 14 علبة من كل مصنع وكما مبين في ادناه، والمطلوب اختبار ان كان هناك فرق جوهري بين كلا المصنعين من ناحية وزن العلب المنتجة عند مستوى معنوية $\alpha = 0.01$.

عينة المصنع الاول :

439, 39 , 47, 43, 47, 40, 39, 51, 45, 50, 50, 43, 48,

عينة المصنع الثاني :

3951, 44, 47 , 49, 42, 38, 52, 49, 45, 51, 46, 38, 52,

الحل لـ **(5.8)** : لدينا :

$s_1=4.4326$, $\bar{x}_1 = 44.428$, $n_1= 14$

$s_2=4.7027$ $\bar{x}_2 = 46.5$, $n_2=14$

وعلى فرض عدم تساوي التباينين نحصل على :

$$t' = \frac{(\bar{x}_1 - \bar{x}_2)-(u_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{44.517 - 45.938 - 0}{\sqrt{\frac{(4.315)^2}{14}+\frac{(5.106)^2}{14}}}$$

$$= \frac{-1.421}{1.78668} = -0.7953$$

وبمقارنة القيمة المحتسبة مع القيمة الجدولية المبينة في الملحق (5.3) وهي : $t_{0.01/2},26 = 2.779$ نجد انها اقل من الجدولية، عليه نقبل فرضية العدم القائلة بعدم وجود فرق جوهري بين متوسطي العينتين .

(1) حالة دراسية رقم C_{8-2}

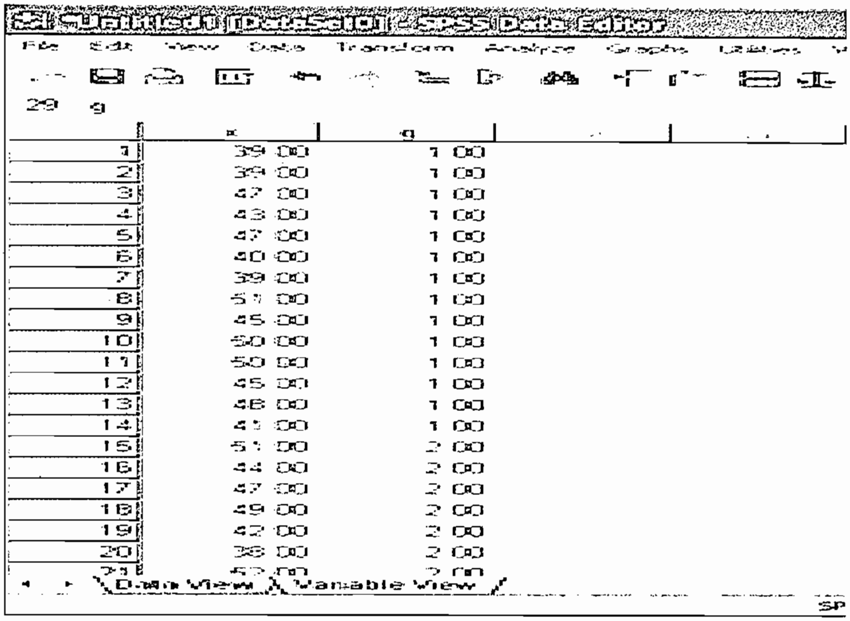
استخدام برنامج SPSS لانجاز اختبارالفروق بين مجتمعين مستقلين بتوظيف معطيات المثال (5.8)

من المفيد الاشارة اولا الى ان صيغة انشاء الملف لهذا الاختبار عند استخدام برنامج SPSS يتطلب ادخال كلا العينتين بذات العمود و في العمود الثاني يتم اعطاء القيمة 1 امام قيم العينة الاولى والقيمة 2 امام قيم العينة الثانية وكما مبين في الشكل البياني رقم (5.8) التالي :

الشكل البياني رقم (5.8)

نموذج انشاء الملف لاختبارالفروق بين مجتمعين مستقلين

Two independent Samples



	x	g
1	39.00	1.00
2	39.00	1.00
3	47.00	1.00
4	43.00	1.00
5	47.00	1.00
6	40.00	1.00
7	39.00	1.00
8	51.00	1.00
9	45.00	1.00
10	50.00	1.00
11	50.00	1.00
12	45.00	1.00
13	48.00	1.00
14	41.00	1.00
15	51.00	2.00
16	44.00	2.00
17	47.00	2.00
18	49.00	2.00
19	42.00	2.00
20	38.00	2.00

وان اجراءات عملية التحليل تتلخص بالخطوات التالية :

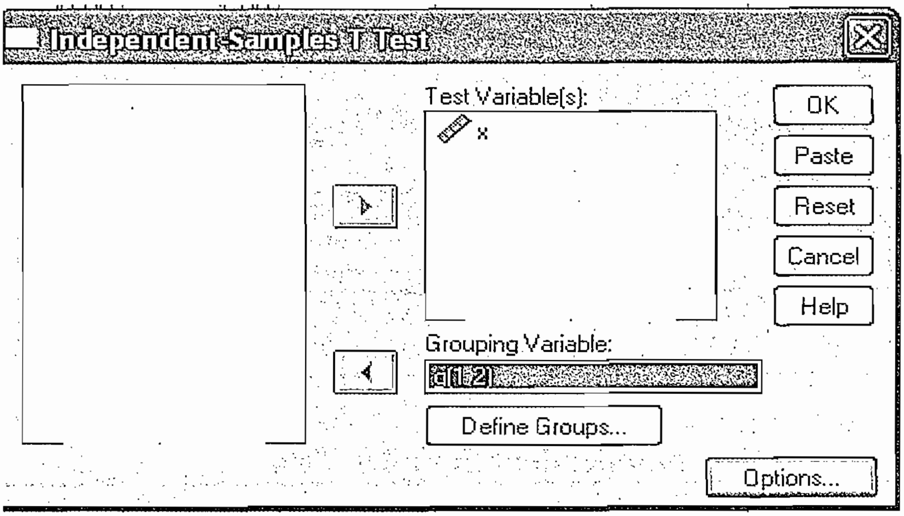
- من قائمة Analysis يتم اختيارالامر الفرعي Compare Mean، ثم الخيار Independent sample T-test ليظهر لنا مربع الحوار المبين في الشكل البياني رقم (6.8)، وفيه يتم نقل المتغير x_i الى الموقع تحت عنوان Test Variable، ونقل المتغير g تحت العنوان Grouping Variable ،

- الكبس على ايقونة Define Group لتظهرلنا الوحة المبينة في الشكل البياني رقم (7.8) ليتم فيها تدوين رموز كل من العينة الاولى والعينة الثانية والتي هي 1 و2 كما مبين في الملف في الشكل البياني رقم (5.8).

- الكبس على ايقونة Continue للعودة الى مربع الحوار، وفي حالة الحاجة لتغير درجة الثقة يتم الكبس على ايقونة Options لاجراء عملية التغير والعودة مرة اخرى الى مربع الحوار للكبس على ايقونة Ok لنحصل على المخرجات المبينة في الجداول رقم (2.8) .

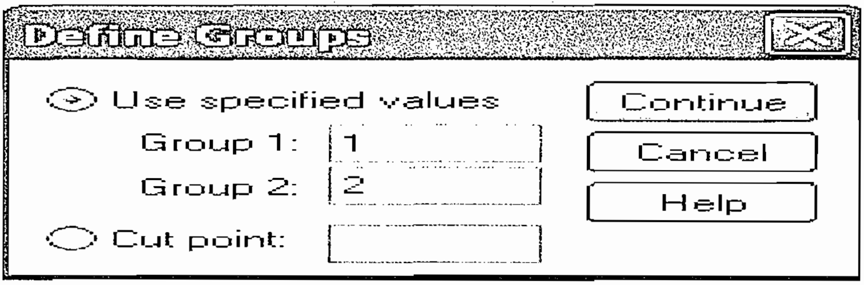
الشكل البياني رقم (6.8)

مربع حوار اختبار الفروق بين متوسطي مجتمعين موزعين طبيعيا



الشكل البيني رقم (7.8)

لوحة تدوين رموز العينات في المتغير g



ومن الجداول رقم (2.8) نجد بان قيمة متوسط هذه الفروق البالغ 1.367- غم يقع ضمن حدي الثقة عند درجة 95 ٪، و ان قيمة t المحتسبة ومقدارها 0.752- هي تقل عن مستوى معنوية 0.05، وعليه نقبل فرضية تماثل اوزان منتجات كلا المصنعين، الاان حصيلة اختبار f غير معنوية مما يشير الى عدم تساوي تبايني العينتين . وقد يعود ذلك لصغر حجم العينة التي يصعب معها تاكيد التوزيع الطبيعي للمجتمع .

جداول رقم (2.8)

يبين مخرجات برنامج SPSS لاختبار T للفروق بين عينتين مستقلتين

Group Statistics

	g	N	Mean	Std. Deviation	Std. Error Mean
x	1.00	14	44.5714	4.41526	1.18003
	2.00	14	45.9286	5.10602	1.36464

Independent Samples Test

		t-test for Equality of Means					
	t	df	Sig. 2-tailed	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						Lower	Upper
x Equal variances assumed	-.752	26	.459	-1.35714	1.80408	-5.06549	2.35120
Equal variances not assumed	-.752	25.469	.459	-1.35714	1.80408	-5.06925	2.35496

جزء من الجدول اعلاه

Independent Samples Test

	Levene's Test for Equality of Variances	
	F	Sig.
x Equal variances assumed	.257	.617
Equal variances not assumed		

(3) اختبار الفرق بين نسبتي مجتمعين

اذا كان $np > 0.05$ او nq فيفترض ان توزيع p طبيعي. فلو فرضنا لدينا مجتمعين ونسب النجاح لهما هي p_1 و p_2، حينئذ سنرمز لنسبة العينة الاولى التي حجمها n_1 بـ p_1 و بـ p_2 لنسبة العينة الثانية التي حجمها n_2 المسحوبة من المجتمع الثاني، وبافتراض استقلالية كلا العينتين، يكون لدينا :

العينة الاولى $\mu_{p_1} = P_1$ و $\sigma_{p_1} = \sqrt{\frac{P_1 Q_1}{n_1}}$ ،

وللعينة الثانية $\mu_{p_2} = P_2$ و $\sigma_{p_2} = \sqrt{\frac{P_2 Q_2}{n_2}}$

وان الفرق بين متوسطي المجتمعين هو : $\mu_{p_1} - \mu_{p_2} = P_1 - P_2$

والخطأ المعياري هو :

$$\sigma_{p_1 - p_2} = \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$$

اما في حالة مساواة $P_1 - P_2$ ، فيتم استبدالها بقيمة مشتركة ولنرمز لها بـ P_c، وبذلك تصبح صيغة الخطا المعياري :

$$\sigma_{p_1-p_2} = \sqrt{\frac{P_cQ_c}{n_1} + \frac{P_cQ_c}{n_2}}$$

$$= \sqrt{p_cq_c\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

حيث ان : $p_c = \dfrac{p_1 + p_2}{n_1 + n_2}$

اما صيغة اختبار الفرق بين نسبتي مجتمعين فتصبح :

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{s_{p1-p2}}$$

مثال (6.8) : لدينا عينتين من العمال من منطقتين وعدد العاطلين بينهم وكالاتي :

$n_1=1600$, $p_1=120$, $n_2=1400$, $p_2=84$. والمطلوب اختبار ان كانت نسبة العاطلين في كلا المنطقتين مختلفة عند مستوى معنوية $\alpha = 0.05$.

الحل لـ (6.8) :

- نحدد الفرضية :

$$H_0 : P_1 - P_2 = 0$$
$$H_1 : P_1 - P_2 \neq 0$$

- يكون في حالتي اكب واقل، فيكون الاختبار من H_1حيث ان قبول جانبين، وبالرجوع الى الملحق رقم (4.2) نجد ان القيمة الجدولية هي :

$$Z_{\alpha/2} = 1.96$$

▫ لدينا :

$$p_c = \frac{p_1 + p_2}{n_1 + n_2} = \frac{120 + 84}{1600 + 1400} = 0.068$$

$$q_c = 1 - p_c = 1 - 0.068 = 0.932$$

$$s_{p1-p2} = \sqrt{p_c q_c \left(\frac{1}{n_1} + \frac{1}{n} \right)} = \sqrt{(0.068)(0.932) \frac{1}{1600} + \frac{1}{1400}} = 0.0092$$

وبتطبيق صيغة الاختبار نحصل على :

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{s_{p1-p2}} = \frac{(0.075 - 0.06) - 0}{0.0092} = 1.63$$

– القرار : حيث ان قيمة Z المحتسبة اقل من قيمة $Z_{\alpha/2} = 1.96$ الجدولية، عليه نقبل فرضية العدم H_0 ، ونستدل على عدم وجود فرق جوهري بين نسبتي العاطلين في كلا المنطقتين .

8- 2- 3 اختبار المقارنات الزوجية Paired Samples T-test

(1) خصائص واجراءات اختبار المقارنات الزوجية

والهدف من استخدامه هو لقياس ظاهرة معينة تحت ظروف مختلفة، كقياس نمو نباتات معينة عند تعرضها للشمس وقياس نموها من دون تعرضها للشمس ومن ثم اختبار ان كان هناك فرق جوهري في نمو هذه النباتات بين كلا الحالتين، او ان يكون القياس قبل تسميد النبتة وبعد التسميد، او بقسمة نوع من النباتات الى قسمين واعطاء كل قسم نوع مختلف من السماد وهكذا.

على عكس الفرضية التي تقوم عليها الاختبارات السابقة، فان الفرضية التي تقوم عليها عملية اختبار المقارنات الزوجية هي ان العينات التي يتم المقارنة بين متوسطاتها غير مستقلة. و يهدف هذا النوع من الاختبار التخلص من اكبر عدد ممكن من العوامل الخارجية التي تؤدي الى التباين بين مجتمعين، من خلال عمل ازواج متشابهه لعدد من المتغيرات. وبدلا من اجراء التحليل لكل قسم من المشاهدات على حده، يجري استخدام الفرق بين كل زوج من المشاهدات واعتباره متغيرا معينا، مفترضين ان هذه الفروق عشوائية ومسحوبة من مجتمع موزعه فروقاته توزيعا طبيعيا. اما صيغة الاختبار فهي :

$$z = \frac{\bar{d} - \mu_d}{\frac{\sigma_d}{\sqrt{n}}}$$

وفي حالة عدم معلومية تباين المجتمع يستعاض بـ $s_{\bar{d}}$ عن σ_d ، ليكون:

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

حيث ان s_d هو الانحراف المعياري لفروقات العينة. عندها تصبح صيغة الاختباركالاتي:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

مثال (7.8) : لدينا عينة تتكون من 10 نباتات ظلية، تم عرضها لمدة ستة اشهرفي موقع يزداد فية الضوء، وامعطيات عن قياس اطوال هذه النباتات قبل وبعد تعرضها للضوء هي كما مبين في الجدول رقم

(3.8) التالي. وامطلوب اختبار ان كان هناك فرق جوهري في اطوالها قبل وبعد تعرضها للاضاءة الاضافية، عند مستوى معنوية $\alpha = 0.01$.

جدول رقم (3.8)

قياسات اطوال 10 نباتات (سم) قبل وبعد تعرضها للضوء الاضافي

الفروق $d_i = x_{i2} - x_{i1}$	الطول بعد التعرض للضوء	الطول قبل التعرض للضوء	رقم المشاهدة
2	33	31	1
-1	32	33	2
1	36	35	3
-1	29	30	4
3	39	36	5
1	38	37	6
0	41	41	7
5	40	35	8
4	43	39	9
2	34	32	10
$\sum d_i = 16$			

الحل لـ (7.8) : لدينا : $\bar{d} = 1.6, s_d = 2.01$

- فرضية الاختبار هي :

$$H_o: \mu_d \geq 0$$

$$H_1: \mu_d \lhd 0$$

- ووفقا للفرضية اعلاه، يكون الاختبار من جانب واحد، اي ان قبول H_1 عندما يكون الفرق اقل، وباستحدام الملحق رقم (5.3) عند مستوى معنوية $\alpha = 0.01$ ودرجات حرية عددها 9 ، فان القيمة الجدولية هي:

$t_{0.01}, 9 = 3.25$

▪ وبتطبيق صيغة الاختبار في حالة مجهولية تباين المجتمع، نحصل على :

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{1.6 - 0}{\frac{2.01}{\sqrt{10}}} = 2.525$$

▪ القرار : حيث ان قيمة $t_{0.01}, 9 = 3.25$ الجدولية هي اكبر من قيمة t المحتسبة، يتم قبول H_0، ونستدل على ان تعرض النباتات للضوء الاضافي خلال ستة اشهر من شانه ان يؤدي الى زيادة جوهرية في اطوالها.

(2) حالة دراسية رقم C_{8-3}

استخدام برنامج SPSS في أختبار المقارنات الزوجية بتوظيف معطيات المثال (7.8)

بعد انشاء ملف بمعطيات المتغيرين المبينة في الجدول (3.8) واخضاعها للتحليل لبرنامج SPSS يتم متابعة الخطوات التالية :

▪ الكبس على خيار Paired samples T-test من الامر الفرعي Compare means من قائمة Analysis ليظهر لنا مربع الحوار المبين في الشكل البياني رقم (8.8)، وفيه يتم استخدام السهم الجانبي لنقل المتغيرين الى تحت عنوان Paired Variables، كذلك اجراء تاشيرلكلا المتغيرين تحت Current Selection مقابل المتغيرات المدونة في مربع الحوار.

▪ واذا لم تكن حاجة لتغير درجة الثقة المقررة وهي 95 ٪، التي تستدعي الكبس على ايقونة Options، عندها يتم الكبس على ايقونة Ok لنحصل على المخرجات المبنة في جداول رقم (4.8).

ومن نتائج التحليل المبينة في الجداول رقم (4.8)، نجد ان التحليل في مرحلته الاولى يعرض متوسطي اطوال النباتات لكل من متغيري قبل وبعد تعرضها للضوء الاضافي وهي :

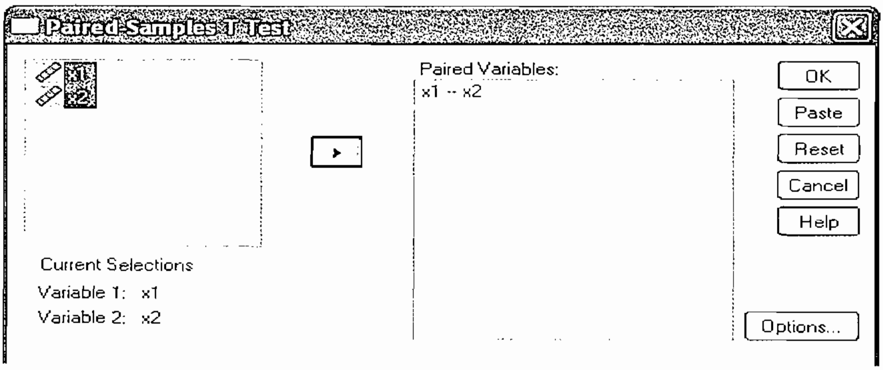
$$\text{Before}\ ,\ \bar{x}_{1i} = 34.9$$
$$\text{After}\ ,\ \bar{x}_{2i} = 36.5$$

وبانحراف معياري مقداره 3.51 و 4.453 على التوالي. وهذا يعني ان تعرض النباتات للضوء الاضافي ادى الى زيادة في النمو بحوالي 1.1 سم، الا ان مقدار الزيادة قد تفاوتت من نبتة لاخرى كما يستدل من الارتفاع الذي طرأ في مقدار الانحراف المعياري، اي ان الضوء الاضافي كان تاثيره متباينا من نبتة لاخرى. وان معامل الارتباط الذي يدل على العلاقة بين الحالتين يشير الى عـــلاقة قوية مقدارها 0.899 وهي معنوية عند 0.000.

اما حصيلة الاختبار فتدل على قبول فرضية العدم عند مـستوى معنويـة 0.05 اي بدرجـة ثقـة مقـدارها % 95، اي ان تعـريض هـذا النـوع مـن النباتات الظلية لضوء اضافي من شانه ان يزيد في معدل نموها بدرجة معقولة

الشكل البياني (8.8)

مربع حوار أختبار المقارنات الزوجية Paired Samples T-test



جداول رقم (4.8)

توضح نتائج اختبار Paired samples T-test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	X1I	34.9000	10	3.5103	1.1101
	X2I	36.5000	10	4.4535	1.4083

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	X1I & X2I	10	.899	.000

Paired Samples Test

	Paired Differences					t	df	ig. (2-taile
				99% Confidence Interval of the Difference				
	Mean	d. Deviatic	Std. Erro Mean	Lower	Upper			
Pair X1I - X	1.6000	2.0111	.6360	1.6081	1.5919	-2.516	9	.033

8- 3 استخدام χ^2 في اختبار التجانس Test of Consistency

8-3-1 خصائص اختبار التجانس والاجراءات

وهو الاختبار الذي يلائم حالة كون العينات مسحوبة من عدة مجتمعات متجانسة وفقا لمعيار التصنيف، ويمكن اجمال خصائص اختبار التجانس بما يلي :

- ان كل العينات المستقلة مسحوبة من مجتمعات معلومة التوزيع مسبقا،
- ان احتساب التكرارات المتوقعة تعتمد على فرضية ان المجتمعات التي تعود اليها العينات هي متجانسة ،

- ان استنتاج التجانس يتعلق بتجانس المجتمعات طبقا لمعيار التصنيف المعني.

اما صيغة الاختبار فهي :

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

8- 3- 2 حالة دراسية رقم C8-5

استخدام برنامج SPSS لانجاز اختبار χ^2 للتجانس

في دراسة قامت بها قناة تلفزيونية لمعرفة كان برنامجها الترفيهي له نفس الاهتمام بين كافة الفئات العمرية، فاختارت عينة من المشاهدين حجمها n = 74 وحصلت على النتائج المبنة في الجدول التالي، والمطلوب استخدام برنامج SPSS لاختبار ان كان هناك فروق في رغبة مشاهدة البرنامج بين الفئات العمرية عند مستوى معنوية $\alpha = 0.05$.

عينة من مشاهدي قناة تلفزيونية حسب

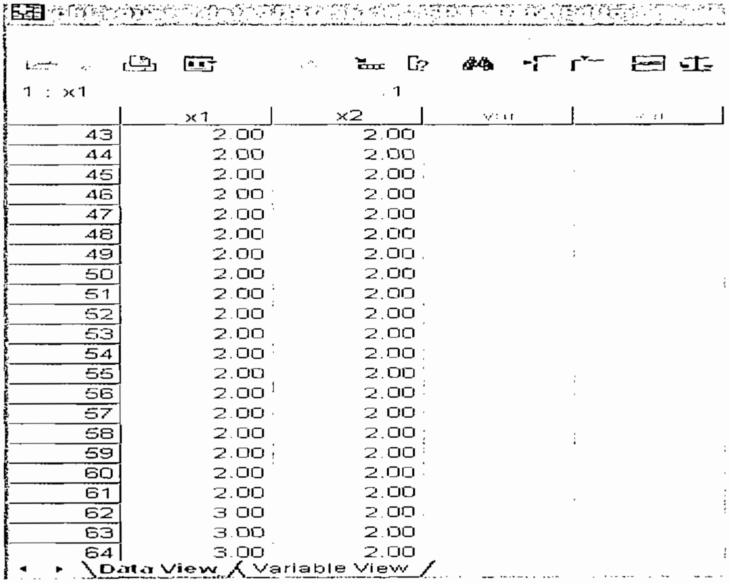
الفئة العمرية والرغبة في مشاهدة البرنامج الترفيهي

الفئة العمرية	مستوى الرغبة			
	لايرغب	يرغب	يرغب جدا	المجموع
أقل من 18	20	16	4	40
18- 50	10	8	3	21
50 فاكثر	6	5	2	13
المجموع	36	29	9	74

- من المفيد الاشارة اولا الى ان تكوين الملف لاستخدام برنامج SPSS في اختبار التجانس يتطلب اعطاء المتغير الاول وهي الفئات العمرية القيم 1 للفئة الاولى والقيم 2 للثانية وتاخذ الفئة الثالثة القيم 3 , وعلى نفس الغرار بالنسبة للمتغير الثاني وهو متغيرالرغبة، تعطى القيم 1 لحالة عدم الرغبة والقيم 2 للرغبة والقيم 3 لحالة راغب جدا، وهذا طبعا لكل قيمة من قيم المعطيات اي لغلية 40 قيمة للفئة الاولى ولغاية 36 قيمة لحالة عدم الرغبة من المتغير الثاني وهكذا، وكما مبين في الشكل البياني رقم (9.8) .

الشكل البياني رقم (9.8)

اسلوب ادخال المعطيات لتكوين ملف لاختبار التجانس



1 : x1

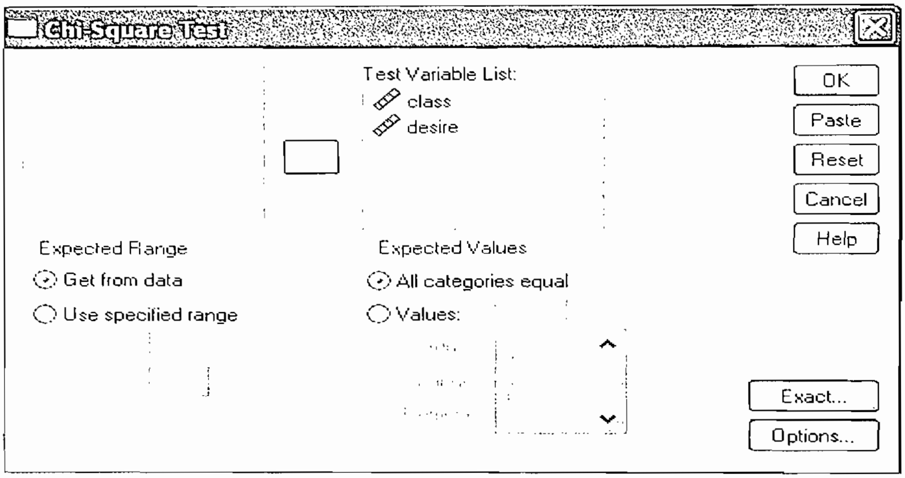
	x1	x2	var	var
43	2.00	2.00		
44	2.00	2.00		
45	2.00	2.00		
46	2.00	2.00		
47	2.00	2.00		
48	2.00	2.00		
49	2.00	2.00		
50	2.00	2.00		
51	2.00	2.00		
52	2.00	2.00		
53	2.00	2.00		
54	2.00	2.00		
55	2.00	2.00		
56	2.00	2.00		
57	2.00	2.00		
58	2.00	2.00		
59	2.00	2.00		
60	2.00	2.00		
61	2.00	2.00		
62	3.00	2.00		
63	3.00	2.00		
64	3.00	2.00		

Data View / Variable View

- يتم اخضاع الملف للامر Analysis ومنه الامر الفرعي Non-parametric test ثم الكبس على الخيار Chi-square ليظهر مربع الحوار المبين في الشكل البياني رقم (10.8) ،

الشكل البياني رقم (10.8)

مربع حوار اختبار التجانس باستخدام مربعات كاي χ^2



- يتم نقل المتغيرين تحت عنوان Test Variable List باستخدام السهم الموجود بجنب مربع الحوار، ومن ثم التاشير عند All Categories Equal،

- الكبس على ايقونة Ok فنحصل على المخرجات المبينة في الجداول رقم (5.8) بضمنها جدول الاحصاء الوصفي الذي يشير الى تشابه قيمتي متوسطي المتغيرين والى تجانس الاراء ضمن الفئات العمرية كما يتضح من قيم الانحراف المعياري لكلا المتغيرين، كما و يستدل على معنوية النتائج عند درجة ثقة 95 ٪ التي جاءت عند درجة معنوية 0.000، حيث ان معنوية asymptotic significance التي تعتمد على توزيع asymp. distribution تعتبر مقبولة عند اقل من 5 ٪. اي قبول H_0 القائلة بتجانس معايير التصنيف لكلا المتغيرين .

جداول رقم (5.8)

مخرجات برنامج SPSS لاستخدام اختبار Chi-Square Test

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
level of wish	74	1.6351	.6939	1.00	3.00

age groups

	Observed N	Expected N	Residual
1.00	40	24.7	15.3
2.00	21	24.7	-3.7
3.00	13	24.7	-11.7
Total	74		

level of wish

	Observed N	Expected N	Residual
1.00	36	24.7	11.3
2.00	29	24.7	4.3
3.00	9	24.7	-15.7
Total	74		

Test Statistics

	age groups	level of wish
Chi-Square[a]	15.595	15.919
df		
	.000	.000

8- 4 تحليل التباين Analysis of Variance

8-4-1 خصائص تحليل التباين والاجراءات

تحليل التباين هو امتداد لاختبارT لاستخدامه في اختبار اكثر من عينتين مع القدرة على تحليل طبيعة ومصدر التباين بين الظواهر المختلفة، حيث يقوم بتقسيم الاختلافات الكلية الى عدة اجزاء لتحديد مصدرها. والفرضيات التي يقوم عليها الاختبار تتلخص بالاتي :

(1) ان العينات عشوائية تعود لمجتمعات موزعة طبيعيا، ويتم التحقق من شرط العشوائية عند سحب العينات،

(2) ان العينات مسحوبة من مجتمعات موزعة طبيعيا، ويتم التحقق من شرط التوزيع الطبيعي باستخدام اختبار χ^2 لاختبار التجانس او الجودة (المطابقة) ،

(3) تساوي تباينات المجتمعات المسحوبة منها العينات، وفي حالة عدم توفر هذا الشرط يتم اللجوء الى استخدام اختبار بارتليت Bartlet او اختبار Hartly ،اي:

$$\sigma^2_1 = \sigma^2_2 = ...= \sigma^2_k = \sigma^2$$

ويتم اجراء اختبار تحليل التباين اعتمادا على الاحصاءة f ونتائجه تنظم بجدول يدعى جدول تحليل التباين. وهناك حالات عديدة يستخدم معها تحليل التباين منها ما هو بمعيار واحد مع عدة مجاميع، ومعيار واحد مع تعدد المستويات في كل مجموعة، ومنها بمعيارين من دون تفاعل داخلي ومعيارين مع تفاعل داخلي وغيرها .

ففي حالة التحليل بمعيار واحد مثلا يتم تصنيف قيم x_i الى k من المجاميع، فعلامات الطلبة تصنف حسب الشعب، وكل شعبة تضم n من الطلاب وعادة ما يشار اليها بالعناصر. ان الاختلاف في قيم X يعزى الى الاختلاف بين القيم الواقعة ضمن المجموعة الواحدة والى الاختلاف بين المجاميع ذاتها. لذلك فان تحليل التباين يستهدف تجزئة التباين الكلي الى جزئين ومن ثم تتم المقارنة بين تبايني الجزئين باستخدام اختبار f ، اذن ما نحتاجه في حالة تحليل التباين بمعيار واحد One-Way Analysis of Variance هو تجزئة مجموع مربعات التباين ودرجات الحرية v الى تباين بين المجموعات Between Groups وتباين ضمن المجموعات Within Groups والذي يدعى احيانا بالبواقي Residuals، أي ان S^2 الذي هو تباين لـ X_{ij} التي هي عناصر المجاميع k هو

$$S^2 = \frac{\sum_{j=1}^{n}\sum_{i=1}^{k}\left(x_{ij} - \bar{x}\right)^2}{kn - 1}$$

حيث ان:

$$\sum_{j=1}^{n}\sum_{i=1}^{k}\left(x_{ij} - \mu_{\bar{x}}\right)^2 = \sum_{j=1}^{n}\sum_{i=1}^{k}\left(x_{ij} - \bar{x}_i\right)^2 + n\sum_{i=1}^{k}\left(\bar{x}_i - \mu_{\bar{x}}\right)^2$$

(مجموع الاختلاف بين المجاميع) (مجموع الاختلاف ضمن المجاميع)

(مجموع الاختلاف (المربعات)

ومن ذلك نستدل انه في حالة ايجاد اي حدين يمكن ايجاد الحد الثالث، فاذا رمزنا لمجموع المربعات الكلي بـ SST ومجموع مربعات الاختلاف بين المجاميع بـ SSB ولمجموع مربعات الاختلاف ضمن الجاميع بـ SSW فان قيم تقديرات متوسط كل منها هو :

▪ متوسط مربعات الاختلاف بين المجاميع :

$$MSB = \frac{n\sum_{i=1}^{k}\left(\bar{x}_i - \mu\right)^2}{k-1}$$

متوسط مربعات الاختلاف ضمن المجاميع :

$$MSW = \frac{\sum_{j=1}^{n}\sum_{i=1}^{k}\left(x_{ij} - \bar{x}_i\right)^2}{n-k}$$

▪ وان صيغة اختبار الفرضية :

$$F = \frac{MSB}{MSW}, F_{k-1, n-k}$$

ويصبح شكل جدول تحليل التباين كالاتي :

اختبارF	متوسط المربعات	مجموع المربعات	درجات الحرية	مصدر التباين
$\frac{MSB}{MSW}$	$\frac{n\sum\left(\bar{x}_i - \bar{x}\right)^2}{k-1}$ $\frac{\sum\sum(x_{ij} - \bar{x}_i)^2}{k(n-1)}$	$n\sum\left(\bar{x}_i - \bar{x}\right)^2$ $\sum\sum\left(\overline{xij} - \bar{x}\right)^2$	k-1 k (n-1)	بين المجاميع SSB ضمن المجاميع (الخطأ العشوائي) SSW
----	---------	$\sum\sum\left(\overline{x_{ij}} - \bar{x}\right)^2$	k (n-1)	الكلي

ويكون القرار هو رفض H_0 اذا كانت قيمة f المحتسبة اكبر من القيمة الجدولية مع درجات حرية $F_{\alpha}, k-1, k(n-1)$.

2-4-8 تحليل التباين بمعيار واحد

One-Way Analysis of Variance

(1) حالة تساوي حجوم العينات

مثال (8.8) : قسمت مدينة عمان الى اربعة مناطق وتم اختيار عينة عشوائية تتكون من 9 مصارف من كل منطقة، واتضح ان عدد المعاملات المصرفية (بالمئات) لكل مصرف اسبوعيا هي كما مبين في الجدول التالي. المطلوب معرفة ان كان هناك فرق جوهري في معدل عدد المعاملات التي تقوم بها المصارف اسبوعيا بين المناطق الاربعة وعند مستوى معنوية $\alpha = 0.05$.

عدد المعاملات المصرفية (بالمئات) لاربعة مناطق في مدينة عمان

المصرف	المناطق			
	X_1	X_2	X_3	X_4
1	5	8	7	10
2	6	7	5	8
3	3	7	6	9
4	2	9	8	9
5	4	10	9	11
6	10	11	10	12
7	7	8	7	9
8	3	4	3	5
9	4	5	4	6

الحل لـ (8.8) : من معطيات الجدول اعلاه لدينا :

$$\sum x_1 = 44, \sum x_2 = 69, \sum x_3 = 39, \sum x_4 = 79, \sum x_i 251$$

- مجموع مربعات الاختلاف بين المناطق (المجاميع) :

$$SSB = n\sum_{i=1}^{4}\left(\bar{x}_i - \mu_{\bar{x}}\right)^2$$
$$= 9\left[(4.89-6.97)^2 + (7.67-6.97)^2 + (6.56-6.97)^2 + (8.78-6.97)^2\right]$$
$$= 74.3454$$

- مجموع مربعات الاختلاف الكلي :

$$SST = n\sum_{i=1}^{9}\sum_{i=1}^{4}\left(x_{ij} - \mu_{\bar{x}}\right)^2$$
$$= 9\left[(5-6.97)^2 + (6-6.97)^2 + \ldots\ldots\ldots + (6-6.97)^2\right]$$
$$= 246.087$$

- مجموع مربعات الاختلاف ضمن المجاميع (المناطق) :

$$SSW = SST - SSB$$
$$= 246.087 - 74.3454 = 171.7422$$

- الفرضية :

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

H_1: على الاقل اثنين من المتوسطات غير متساوية

وفي ضوء النتائج اعلاه، نحصل على جدول تحليل التباين التالي :

مصدر التباين	درجات الحرية	مجموع المربعات	متوسط المربعات	*f*
بين المجاميع SSB	k-1=3	74.3454	24.7818	4.6175
ضمن المجاميع SSW	K(n-1)=32	171.7422	5.3669	
المجموع الكلي	nk-1=35	246.0876	-----	-----

- القرار : باستخدام الملحق رقم (7.3) نجد ان قيمة *f* الجدولية هي: $f_{0.05,} 3{,}32 = 5.239$، وحيث ان القيمة المحتسبة اقل من القيمة الجدولية، عليه نقبل H_0 ونستدل على عدم وجود فروق جوهرية بين متوسطات المناطق .

مع الاشارة الى انه بالامكان اختصار عمليات مجاميع المربعات من خلال استخدام الصيغ التالية، المشتقة من الصيغ اعلاه :

$$SST = \sum_{i=1}^{k}\sum_{i=1}^{n} x_{ij}^2 - \frac{\left(\sum_{i=1}^{k}\sum_{i=1}^{n} x_{ij}\right)^2}{kn}$$

$$SSB = \frac{\sum_{i=1}^{k}\left(\sum_{i=1}^{n} x_{ij}\right)^2}{n} - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}\right)^2}{kn}$$

$$SSW = SST - SSB$$

(2) تحليل التباين بمعيار واحد في حالة عدم تساوي حجوم العينات

وفيها يتم اتباع نفس الاجراءات، باستثناء اجراء تعديل بسيط وهو اعتبارحجم العينة يساوي n_i بدلا من n، اي ان مجموع العناصر $n_i = n_1 + n_2 + ... + n_k$ بدلا من nk، وبذلك تكون صيغ الحساب كالاتي:

$$SSB = \sum_{i=1}^{k} \frac{\left(\sum_{i=1}^{n_i} x\right)}{n_i} - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}\right)^2}{\sum n_i}$$

$$SST = \sum_{i=1}^{k}\sum_{i=1}^{n_i} x_{ij}^2 - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}\right)^2}{\sum n_i}$$

حيث ان : $n = \sum n_i$

مثال (9.8) : لنفرض لدينا اربعة مجاميع (عينات)، وان عدد عناصر كل مجموعة يختلف عن الاخرى، وكما مبين في الجدول التالي، والمطلوب اختبار فرضية من ان متوسطات المجتمعات المسحوبة منها العينات متساوية : $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ عند مستوى معنوية $\alpha = 0.05$.

المجاميع			
4	3	2	1
37	35	35	38
34	36	35	37
34	36	36	36
37	36	37	37
37	37	34	37
36	37	34	36
	36	37	37
	34	35	38
	34	34	
	36	36	
	35		
	35		
	35		
$n_4=6$	$n_3=13$	$n_2=10$	$n_1=8$
$\sum_{i=1}^{6} x_i = 215$		$\sum_{i=1}^{13_i} x_i = 462$ $\sum_{i=1}^{10_i} x_i = 353$	$\sum_{i=1}^{8_i} x_i = 296$

الحل لـ (9.8) :

- الفرضية :

$$H_0 : \mu_1 = \mu_2 = \mu$$
$$H_1 : \mu_1 \neq \mu_2 \neq \mu$$

لدينا :

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} = 38 + 37 + ... + 37 + 36 = 1326$$

- مجموع العناصر

مجموع مربعات العناصر:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} x^2 = (38)^2 + (37)^2 + \dots\dots\dots + (37)^2 + (36)^2 = 47574$$

- مجموع مربعات العناصر مقسوما على حجم العينة :

$$\sum_{i=1}^{k} \frac{\left(\sum_{j=1}^{n_i} x\right)^2}{n_i} = \frac{(296)^2}{8} + \frac{(353)^2}{10} + \frac{(462)^2}{13} + \frac{(215)^2}{6}$$

$$= 47535.8$$

- معامل التصحيح : $\frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x\right)^2}{\sum n_i} = \frac{(1326)^2}{37} = 47521$

يكون لدينا :

- مجموع مربعات التباين :

$$SST = \sum_{i=1}^{k}\sum_{i=1}^{n_i} x_{ij}^2 - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}\right)^2}{\sum n_i} = 47574 - 47521 = 53$$

- مربعات التباين بين المجموعات :

$$SSB = \sum_{i=1}^{k} \frac{\left(\sum_{i=1}^{n_i} x\right)}{n_i} - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}\right)^2}{\sum n_i} =$$

$$47535.8 - 47521 = 14.8$$

- مربعات التباين ضمن المجموعات :

$$SSW = SST - SSB = 53 - 14.8 = 38.2$$

- ومن النتائج اعلاه نحصل على جدول تحليل التباين التالي :

مصدر التباين	درجات الحرية df	مجموع المربعات SS	متوسط المربعات μ_{ss}	f
بين المجموعات SSB	k-1=3	14.8	4.93	$\frac{4.93}{1.16} = 4.25$
ضمن المجموعات SSW	n_i-a=33	38.2	1.16	
مجموع التباين SST	$\sum_{i=1}^{k} n_i - a = 36$	53	1.61	

- القرار : عند درجات حرية 3 و33، ومستوى معنوية 0.05 نجد ان قيمة *f* الجدولية (من الملحق7.3) هي 5.462، وحيث ان *f* المحتسبة هي اقل من الجدولية، عليه نقبل فرضية العدم H_0 ونستدل على عدم وجود فروق جوهرية بين المتوسطات .

(3) حالة دراسية رقم C_{8-5}

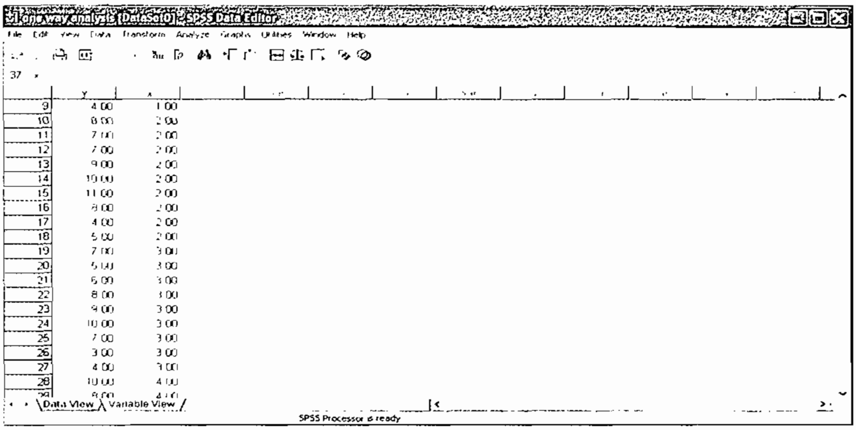
استخدام برنامج SPSS لانجاز تحليل التباين بمعيار واحد بتوظيف معطيات المثال رقم (8.8)

- لكون لدينا متغير بعدة مستويات (مجاميع)، وعليه نستخدم تحليل التباين بمعيار واحد One-Way Analysis of Variance وأول خطوة مطلوبة في استخدام برنامج SPSS هي اعداد ملف المعطيات بوضع قيم كافة المناطق في متغير (عمود) واحد كمتغير تابع، ووضع رموز كل منطقة امام قيمها الواردة في المتغير التابع لتشكيل المتغير المستقل او ما يدعى Factor وكما مبين في الشكـل البياني رقم (11.8)،

الشكل البياني رقم (11.8)

يبين شكل ملف المدخلات لتحليل التباين بمعيار واحد

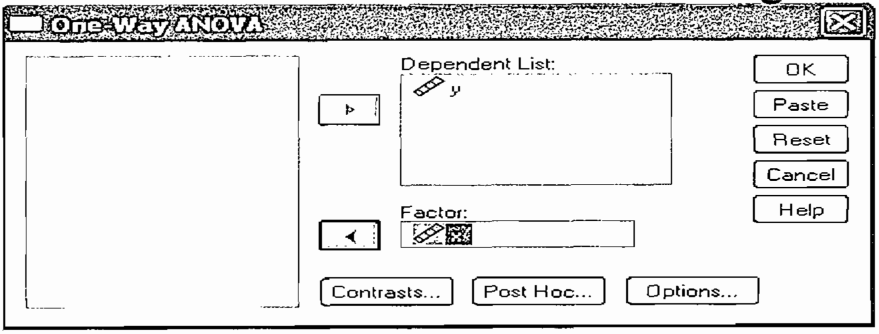
One-Way ANOVA



▪ استخدام قائمة Analysis ومنها الامر الفرعي Compare mean ومن ثم الخيار One-Way Analysis of Variance فيظهر لنا مربع الحوار المبين في الشكل البياني رقم (12.8)، وفيه يتم تحويل المتغير التابع الى خانة Dependent List، والمتغير المستقل الى خانة Factor باستخدام السهم الجانبي الموجود في مربع الحوار، وكما موضح على الشكل البياني المذكور.

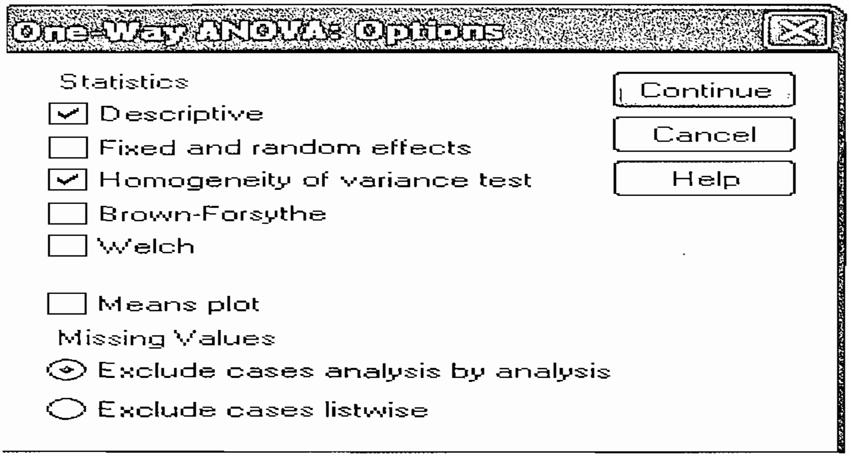
الشكل البياني رقم (12.8)

مربع الحوار لتحليل التباين بمعيار واحد One-Way ANOVA



- الكبس على ايقونة Options الموجودة في مربع الحوار اعلاه، فتظهر لنا لوحة الخيارات المبينة في الشكل البياني رقم (13.8)، ليتم التاشير على ماهو مطلوب منها مثل المقاييس الوصفية Descriptive والتحقق من تجانس التباين Homogeneity of Variance وما الى ذلك. وبعد الانتهاء من تحديد الخيارات يتم الكبس على ايقونة Continue للعودة الى مربع الحوار من جديد .

الشكل البياني رقم (13.8) لوحة خيارات ايقونة Options

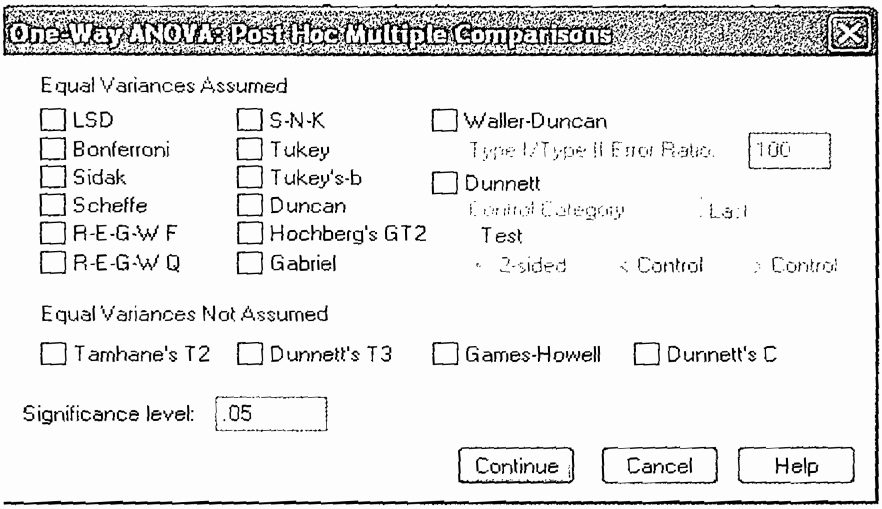


- القيام بالكبس على ايقونة Post Hoc الموجودة في مربع الحوار ايضا، فتظهر لنا اللوحة المبينة في الشكل البياني رقم (14.8)، والخيارات التي توفرها اللوحة المذكورة تتعلق باختيار طريقة اختبار فرضية تساوي التباينات Equal Variance Assumed كان تكون طريقة Turkey ، وفرضية عدم التساوي Equal Variance Not Assumed كاختار طريقة Dunnett's C مثلا، بالاضافة الى مستوى المعنوية Level of Significance المطلوبة ان كانت تختلف عن 0.05 . وحال الانتهاء من تحديد الخيارات يتم الكبس على ايقونة Continue

للعودة الى مربع الحوار مرة اخرى، وفيه يتم الكبس على ايقونة Ok لنحصل على مخرجات التحليل المبينة في الجداول رقم (6.8) .

الشكل البياني رقم (14.8)

لوحة الخيارات البعدية Post Hoc وهي طريقة التحقق من فرضية التجانس وعدم التجانس بين التباينات وتحديد مستوى المعنوية



- ومن المخرجات المبينة في الجداول رقم (6.8) نلاحظ التماثل في النتائج مع ما تم الحصول عليه عند حل المثال (8.8) يدويا، اي ليس هناك فروقا جوهرية واضحة بين متوسطات عدد المعاملات المصرفية بين المناطق خاصة بين المناطق 1 و 2 و 4 كما يتضح من جدول Descriptives مما انعكس ايجابا على معنوية F عند درجة ثقة 95 ٪ مع درجات حرية 3 و 32، فجاءت عند Sig. 0.008 كما يتضح من جدول ANOVA، الامر الذي يقودنا الى قبول فرضية العدم H_0 ورفض الفرضية البديلة H_1 القائلة بعدم التجانس في حجم النشاط

المصرفي بين المناطق الاربعة، وهو ماتؤكده ايضا فترات الثقة المعنوية لهذه المتوسطات كما هو مبين من جدول Multiple Comparisons. بالاضافة الى ما تشير اليه المخرجات الى التجانس بين تباينات هذه المناطق كما يتبين من جدول Descriptives ايضا .

جداول رقم (6.8)

يبين مخرجات برنامج SPSS في تحليل التباين

Descriptives

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1.00	9	4.8889	2.47207	.82402	2.9887	6.7891
2.00	9	7.6667	2.23607	.74536	5.9479	9.3855
3.00	9	6.5556	2.29734	.76578	4.7897	8.3214
4.00	9	8.7778	2.22361	.74120	7.0686	10.4870
Total	36	6.9722	2.64560	.44093	6.0771	7.8674

Test of Homogeneity of Variances

Levene Statistic	df1	df2	Sig.
.062	3	32	.979

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	74.306	3	24.769	4.644	.008
Within Groups	170.667	32	5.333		
Total	244.972	35			

Multiple Comparisons

Dependent Variable: y

	(I) x	(J) x	Mean Difference (I-J)	Std. Error	Sig.
Tukey HSD	1.00	2.00	-2.77778	1.08866	.071
		3.00	-1.66667	1.08866	.432
		4.00	-3.88889(*)	1.08866	.006
	2.00	1.00	2.77778	1.08866	.071
		3.00	1.11111	1.08866	.739
		4.00	-1.11111	1.08866	.739
	3.00	1.00	1.66667	1.08866	.432
		2.00	-1.11111	1.08866	.739
		4.00	-2.22222	1.08866	.194
	4.00	1.00	3.88889(*)	1.08866	.006
		2.00	1.11111	1.08866	.739
		3.00	2.22222	1.08866	.194
Dunnett T3	1.00	2.00	-2.77778	1.11111	.125
		3.00	-1.66667	1.12491	.605
		4.00	-3.88889(*)	1.10833	.017
	2.00	1.00	2.77778	1.11111	.125
		3.00	1.11111	1.06863	.872
		4.00	-1.11111	1.05116	.863
	3.00	1.00	1.66667	1.12491	.605
		2.00	-1.11111	1.06863	.872
		4.00	-2.22222	1.06574	.259
	4.00	1.00	3.88889(*)	1.10833	.017
		2.00	1.11111	1.05116	.863
		3.00	2.22222	1.06574	.259

- The mean difference is significant at the .05 level.

	(I) x	(J) x	Mean Difference (I-J)	95% Confidence Interval	
				Lower Bound	Upper Bound
Tukey HSD	1.00	2.00	-2.77778	-5.7274	.1718
		3.00	-1.66667	-4.6162	1.2829
		4.00	-3.88889(*)	-6.8385	-.9393
	2.00	1.00	2.77778	-.1718	5.7274
		3.00	1.11111	-1.8385	4.0607
		4.00	-1.11111	-4.0607	1.8385
	3.00	1.00	1.66667	-1.2829	4.6162
		2.00	-1.11111	-4.0607	1.8385
		4.00	-2.22222	-5.1718	.7274
	4.00	1.00	3.88889(*)	.9393	6.8385
		2.00	1.11111	-1.8385	4.0607
		3.00	2.22222	-.7274	5.1718
Dunnett T3	1.00	2.00	-2.77778	-6.0812	.5257
		3.00	-1.66667	-5.0092	1.6759
		4.00	-3.88889(*)	-7.1845	-.5933
	2.00	1.00	2.77778	-.5257	6.0812
		3.00	1.11111	-2.0624	4.2846
		4.00	-1.11111	-4.2325	2.0103
	3.00	1.00	1.66667	-1.6759	5.0092
		2.00	-1.11111	-4.2846	2.0624
		4.00	-2.22222	-5.3873	.9428
	4.00	1.00	3.88889(*)	.5933	7.1845
		2.00	1.11111	-2.0103	4.2325
		3.00	2.22222	-.9428	5.3873

Y

	x	N	Subset for alpha = .05	
			1	2
Tukey HSD(a)	1.00	9	4.8889	
	3.00	9	6.5556	6.5556
	2.00	9	7.6667	7.6667
	4.00	9		8.7778
	Sig.		.071	.194

Means for groups in homogeneous subsets are displayed.

a Uses Harmonic Mean Sample Size = 9.000.

8- 4- 3 تحليل التباين بمعيار واحد مع اكثر من مستوى واحد للمجموعة الواحدة Nested Analysis of Variance

لدينا k ترمز الى عدد مجاميع الظاهرة، n ترمز لحجم العينة ، m ترمز لعدد المستويات. فتصبح صيغ تحليل التباين لمعيار واحد مع مستويين فاكثر على الشكل التالي :

- مجموع مربعات التباين (الاختلاف) الكلي :

$$SST = \sum^{k}\sum^{m}\sum^{n} x_i^2 - \frac{\left(\sum^{k}\sum^{m}\sum^{n} x_i\right)^2}{kmn}$$

- مجموع مربعات التباين (الاختلاف) بين المجاميع :

$$SSB = \frac{\sum^{k}\left(\sum^{m}\sum^{n} x_i\right)^2}{mn} - \frac{\left(\sum^{k}\sum^{m}\sum^{n} x_i\right)^2}{kmn}$$

- مجموع مربعات التباين (الاختلاف) بين المجاميع الجزئية :

$$SSSB = \frac{\sum^{k}\sum^{m}\left(\sum^{n} x_i\right)}{n} - \frac{\sum^{k}\left(\sum^{m}\sum^{n} x_i\right)^2}{nm}$$

- مجموع مربعات التباين (الاختلاف) ضمن المجاميع الجزئية :

$$SSSW = \sum^{k}\sum^{m}\sum^{n} x_i^2 - \frac{\sum^{k}\sum^{m}\left(\sum^{n} x_i\right)^2}{n}$$

فيكون شكل جدول تحليل التباين لمعيار واحد ولعدة مستويات التالي :

مصدر التباين	درجات الحرية d.f.	مجموع المربعات SS	متوسط المربعات MS	f
بين المجاميع	k-1	SSB	$\frac{SSB}{k-i}$	$\frac{MSSB}{MSSSB}$
بين المجاميع الجزئية	k(m-1)	SSSB	$\frac{SSSB}{k(m-1)}$	$\frac{MSSSB}{MSSSW}$
ضمن المجاميع الجزئية	km(n-1)	SSSW	$\frac{SSSW}{km(n-1)}$	
المجموع الكلي	kmn-1	SST	------	

مثال (10.8) : في الجدول التالي اوزان (كغم) لانتاج احدى انواع اشجار الفاكهة لسنتين n = 2 ، وفي كل سنة اخذت اربعة اشجار m = 4 ، من ثلاثة حقول k = 3 . المطلوب اختبار ان كانت هناك فروق معنوية في متوسط انتاجية الاشجار في السنتين بين هذه الحقول.

المجاميع												
k = 3	الحقل الاول				الحقل الثاني				الحقل الثالث			
m = 4	1	2	3	4	1	2	3	4	1	2	3	4
n = 2	58.5	77.8	84.0	70.1	69.8	56.0	50.7	63.8	56.6	77.8	69.9	62.1
	59.5	80.9	83.0	68.3	69.8	54.5	49.3	65.8	57.5	79.2	69.2	64.5
$\sum^{n} x_i$	118.0	158.7	167.6	138.4	139.6	110.5	100.0	129.6	114.1	157.0	139.1	126.6
$\sum^{m}\sum^{n} x_i$	582.7				479.7				536.8			

الحل لـ (10.8) :

لدينا :

- المجموع الاجمالي :

$$\sum^{k}\sum^{m}\sum^{n} x_i = 582.7 + 479.7 + 536.8 = 1599.2$$

- مجموع مربعات العناصر :

$$\sum^{k}\sum^{m}\sum^{n} x_i^2 = (58.5)^2 + (77.8)^2 + \dots\dots + (64.5)^2 = 108962$$

- مجموع مربعات المجاميع مقسومة على عدد المستويات n :

$$\frac{\sum^{k}\sum^{m}\left(\sum^{n} x_i\right)^2}{n}$$

$$= \frac{(118.0)^2 + (158.7)^2 + \dots\dots + (126.6)^2}{2} = 108946.38$$

- مجموع مربعات المجاميع مقسومة على عينة المجاميع mn :

$$\frac{\sum^{k}\left(\sum^{m}\sum^{n}x_i\right)^2}{mn}$$

$$=\frac{(582.7)^2+(479.7)^2+(536.8)^2}{(2)(4)}=107225.7$$

- مربع المجموع الكلي مقسوما على مجموع عدد الخلايا kmn :

$$\frac{\left(\sum^{k}\sum^{m}\sum^{n}x_i\right)^2}{kmn}=\frac{(1599.2)^2}{(3)(4)(2)}=106560.026$$

فيكون لدينا :

- مجموع مربعات التباين الكلي SST

$$SST=\sum^{k}\sum^{m}\sum^{n}x_i^2-\frac{\left(\sum^{k}\sum^{m}\sum^{n}x_i\right)^2}{kmn}$$

$$=108692-106560.026=2401.973$$

- مجم

وع مربعات التباين بين المجاميع SSB

$$SSB=\frac{\sum^{k}\left(\sum^{m}\sum^{n}x_i\right)^2}{mn}-\frac{\left(\sum^{k}\sum^{m}\sum^{n}x_i\right)^2}{kmn}$$

$$=107225.702-106560.026=665.6758$$

- مجموع مربعات التباين بين المجاميع الجزئية SSSB

$$SSSB=\frac{\sum^{k}\sum^{m}\left(\sum^{n}x_i\right)}{n}-\frac{\sum^{k}\left(\sum^{m}\sum^{n}x_i\right)^2}{nm}$$

$$=108946.38-107225.703=1720.68$$

▪ مجموع مربعات التباين ضمن المجاميع الجزئية SSSW

$$SSSW=\sum^{k}\sum^{m}\sum^{n}x_i^2-\frac{\sum^{k}\sum^{m}\left(\sum^{n}x_i\right)^2}{n}$$

$$=108962.2-108946.38=15.62$$

ومن نتائج العمليات الحسابية اعلاه، نحصل على جدول تحليل التباين التالي :

مصدر التباين	درجات الحرية d.f.	مجموع المربعات SS	متوسط المربعات MS	f
بين المجاميع	2	665.6759	332.838	f_1=1.741 f_2=146.88
بين المجاميع الجزئية	9	1720.6775	191.1864	
ضمن المجاميع الجزئية	12	15.62	1.3017	
المجموع الكلي	23	2401.9734	------	-------

▪ **القرار** : باستخدام الملحق رقم (7.3) وعند مستوي معنوية 0.05 نجد ان القيم الجدولية هي :

$$f_{1,0.05,(2,9)}=10.11$$

$$f_{2,0.05,(9,12)}=4.906$$

وحيث ان القيمة المحتسبة لـ f_1 هي اقل من الفيمة الجدولية، عليه نقبل H_0 نستدل على عدم وجود فروق جوهرية بين المجاميع، في حين نرفض H_0 ونستدل على وجود فروق جوهرية ضمن المجاميع (الحقول) كما يتضح من مقارنة f_2 المحتسبة مع الجدولية .

8- 4- 4 تحليل التباين بمعيابن

Two Ways Analysis of Variance

(1) خصائص واجراءات تحليل التباين بمعيابن

ويهدف الى دراسة تاثيرعاملين (معيارين للتصنيف) على ظاهرة ما (المتغير التابع)، كأن يكون معيار الطلبة ومعيار طرق التدريس مثلا، وكل منهما يضم عدة مستويات او تقسيمات، للوقوف على معرفة تاثير كل من المعيارين الاول والثاني .

وتحليل التباين بمعيارين ممكن ان يتم اما من دون تفاعل داخلي without Internal Interaction، والافتراض يتضمن بان العاملين (المعيارين) لايتفاعلان معا في التاثير على المتغير التابع، اي ان تاثير الاعمدة هو ذاته مع كل صنف او عامل، عندها يطلق عليه تحليل التباين بمعيارين من دون تفاعل داخلي. وفيه يقسم مجموع المربعات الكلي SST :

$$SST = \sum^{k}\sum^{n} x_{ij}^{2} - \frac{\left(\sum^{k}\sum^{n} x_{ij}\right)^{2}}{kn}$$

الى ثلاثة مركبات هي :

- مجموع مربعات التباين بين الصفوف :

$$SSR = \frac{\sum^{n}\left(\sum^{k} x_{i}\right)^{2}}{k} - \frac{\left(\sum^{k}\sum^{n} x_{ij}\right)^{2}}{kn}$$

- مجموع مربعات التباين بين الاعمدة :

$$SSC = \frac{\sum^{k}\left(\sum^{n} x_{j}\right)^{2}}{n} - \frac{\left(\sum^{k}\sum^{n} x_{ij}\right)^{2}}{kn}$$

- مجموع مربعات تباين الاخطاء (ضمن الاعمدة)

$$SSE = SST - (SSR + SSC)$$

مثال (11.8) : المعطيات في الجدول التالي تمثل نتائج تجربة زراعية تهدف معرفة تاثير 4 اصناف من الحنطة، و 3 انواع من الاسمدة في زيادة متوسط انتاجية الدونم الواحد من الحنطة. المطلوب اختبار ان كانت هناك فروق جوهرية بين متوسطات انتاجية الدونم الواحد من اصناف الحنطة، وكذلك بين متوسطات انتاجية الدونم الواحد باختلاف نوع السماد تحت مستوى $\alpha_1 = 0.01$ و $\alpha_2 = 0.05$.

نوع السماد	صنف الحنطة (القمح)				المجموع
	a	b	c	d	
1	10	7	8	5	20
2	9	7	5	4	25
3	8	6	4	4	22
المجموع	27	20	17	13	77

الحل لـ (11.8) :

- نحدد الفرضية :

$$H_0 : \mu_a = \mu_b = \mu_c = \mu_d$$
$$H_0 : \mu_1 = \mu_2 = \mu_3$$

- مجموع المربعات الكلي SST :

- مجموع مربعات التباين بين الصفوف SSR :

$$SST = \sum^{k}\sum^{n} x_{ij}^2 - \frac{\left(\sum^{k}\sum^{n} x_{ij}\right)^2}{kn}$$

$$= (10)^2 + (7)^2 + \dots\dots\dots + (4)^2 - \frac{(77)^2}{12} = 46.92$$

$$SSR = \frac{\sum^{n}\left(\sum^{k} x_i\right)^2}{k} - \frac{\left(\sum^{k}\sum^{n} x_{ij}\right)^2}{kn}$$

$$= \frac{\left[(30)^2 + (25)^2 + (22)^2\right]}{4} - \frac{(77)^2}{12} = 8.17$$

- مجموع مربعات التباين بين الاعمدة SSC :

$$SSC = \frac{\sum^{k}\left(\sum^{n} x_j\right)^2}{n} - \frac{\left(\sum^{k}\sum^{n} x_{ij}\right)^2}{kn}$$

$$= \frac{\left[(27)^2 + (20)^2 + (17)^2 + (13)^2\right]}{n} - \frac{(17)^2}{12} = 34.92$$

- مجموع مربعات تباين الاخطاء (ضمن الاعمدة) SSE :

$$SSE = SST - (SSR + SSC)$$

$$= 46.92 - 8.17 - 34.92 = 3.83$$

وبترتيب النتائج اعلاه نحصل على جدول تحليل التباين التالي :

f	متوسط المربعات MS	مجموع المربعات SS	درجات الحرية d.f.	مصدر التباين
$\frac{11.64}{0.64} = 18.19$	**11.64**	**34.92**	**3**	**بين الاعمدة (الاصناف) SSC**
$\frac{4.09}{0.64} = 6.39$	**4.09**	**8.17**	**2**	**بين الصفوف (الاسمدة) SSR**
	0.64	**3.83**	**6**	**الخطأ (ضمن الاعمدة) SSE**
------	-----	**46.92**	**11**	**المجموع الكلي**

- **القرار** : بالرجوع الى الملحق رقم (7.3)، نجد ان قيمة f الجدولية هي :

$$f_{0.05;3,6} = 4.76$$
$$f_{0.05;2,5} = 5.14$$

ومن خلال المقارنة نستدل على رفض H_0 مما يدل على عدم تساوي متوسطات انتاجية اصناف القمح سواء عنسد 0.05، وكذلك على نطاق نوع السماد، حيث القيم المحتسبة هي اكبر من القيم الجدولية .

او **الافتراض بوجود تفاعل بين العاملين،** عندها يقسم مجموع مربعات التباين الكلي الى 4 مركبات هي : مركبتي العاملين الاول والثاني، والثالثة للتفاعل بين العاملين الاول والثاني، والمركبة الرابعة للخطأ e_{ijk} التي تكون مغيرا مستقلا يتبع التوزيع الطبيعي $N(0,1)$، وان :

$$i = 1,2,\ldots\ldots,r$$
$$j = 1,2,\ldots\ldots,c$$
$$k = 1,2,\ldots\ldots,n$$

حيث ان : r ترمز اى عدد الصفوف، و c تشير الى عدد الاعمدة، و n عدد مشاهدات كل خلية .

وكذلك اثر التفاعل Interaction بين هذين العاملين على المتغير التابع، وذلك لاختبار فرضية تساوي متوسط المتغير التابع مع متوسطات مستويات العوامل، مقابل فرضية عدم وجود تفاعل بين العاملين. وكذا الاجراءات في حالة 3- Way ANOVA مع استخدام برنامج SPSS .

وجميع حالات الاختبار تتم على اساس استيفاء الشروط التي سبق تناولها والمتعلقة بتوزيع المتغير التابع توزيعا طبيعيا، وتساوي التباين، واستقلالية المشاهدات عن بعضها. كما ان التحليل باستخدام برنامج SPSS ممكن ان يتم بكلتا الحالتين بدون او مع وجود تفاعل داخلي بمجرد الاشارة على الخيارالمطلوب على لوحة Univariate : Model المبينة في

الشكل البياني رقم (20.7) وكما سيتضح عند استخدام برنامج SPSS في عملية التحليل في الفقرة التالية .

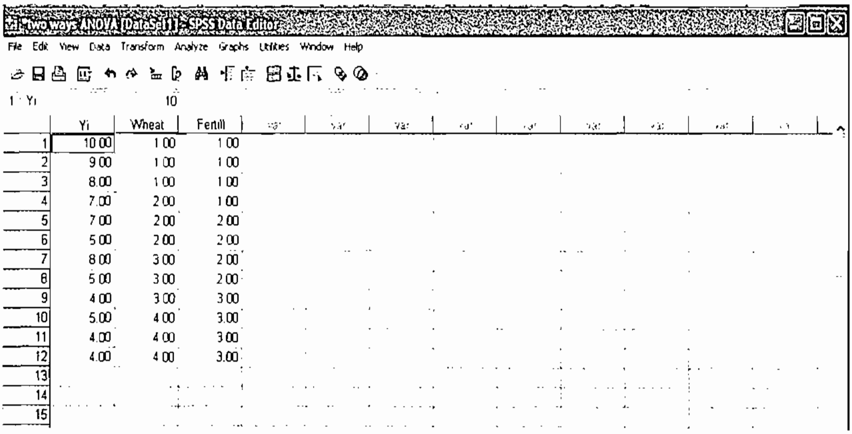
(2) حالة دراسية رقم C_{8-6}

استخدام برنامج SPSS لتحليل التباين بمعيارين
بتوظيف معطيات المثال (11.8)

- تهيئة ملف المعطيات، باعطاء الرموز للعوامل وكذلك لاصناف كل عامل (معيار)، فبالنسبة للمثال (11.8) الذي سيتم اخضاعه للتحليل هنا، فان قيم الخلايا لجدول المعطيات تم الرمز لها بـ y_i كمتغير تابع Dependent Variable، والرموز من 1، 2، 3، 4 لاصناف العامل الاول وهو القمح Wheat، ولاصناف العامل الثاني وهو السماد fertill الرموز 1، 2، 3، كمتغيرات مستقلة Factors، وبذلك يكون شكل الملف لدينا كما مبين في الشكل البياني رقم (15.8).

الشكل البياني رقم (15.8)

شكل ملف تحليل التباين بمعيارين Two Ways Analysis of Variance

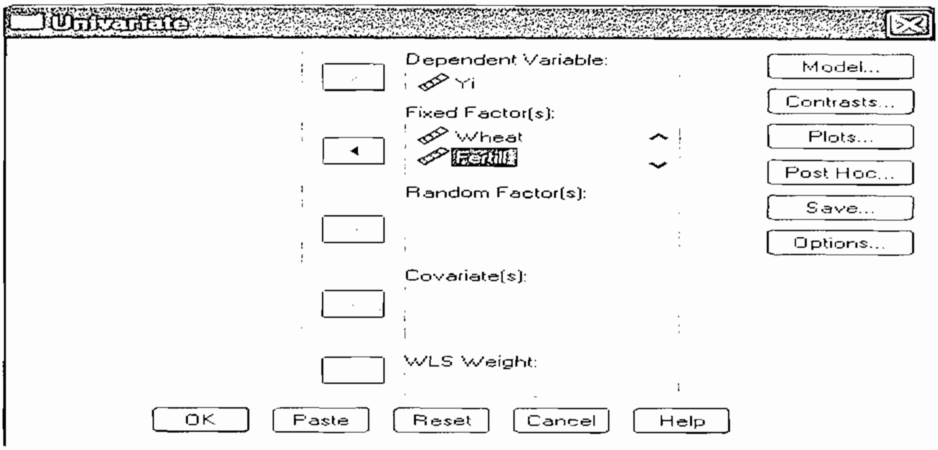


	Yi	Wheat	Fertill
1	10.00	1.00	1.00
2	9.00	1.00	1.00
3	8.00	1.00	1.00
4	7.00	2.00	1.00
5	7.00	2.00	2.00
6	5.00	2.00	2.00
7	8.00	3.00	2.00
8	5.00	3.00	2.00
9	4.00	3.00	3.00
10	5.00	4.00	3.00
11	4.00	4.00	3.00
12	4.00	4.00	3.00
13			
14			
15			

- من قائمة Analysis نختار الامر الفرعي General Linear Model ومنه الخيار Univariate فنحصل على مربع الحوار Univariate المبين في الشكل البياني رقم (16.8)، وفيه يتم استخدام الاسهم الجانبية لنقل المتغير y_i الى النافذة التي تحت Dependent Variable، والمتغيرين Wheat و fertill الى النافذة التي تحت Factors Fixed،

الشكل البياني رقم (16.8)

مربع الحوار Univariate تحليل التباين بمعيارين



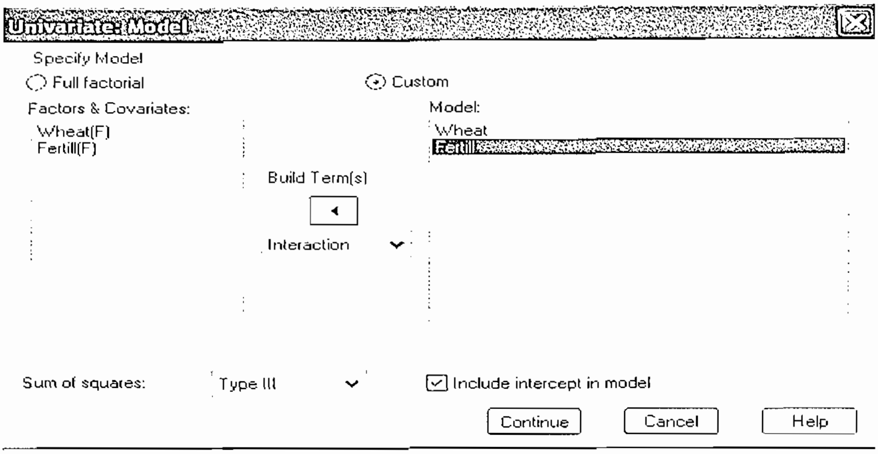
- الكبس على ايقونة Model فتظهر اللوحة Univariate : Model المبينة في الشكل البياني رقم (17.8)، وفيها يتم اجراء التالي :

- التاشير عند Custom للتحكم بالعوامل والتفاعلات وحسب متطلبات التحليل
- الكبس على السهم ذات الاتجاه السفلي الموجود في الوسط تحت Build Term لاختيار Main effects .
- نقل المتغيرين من النافذة التي تحت Factors & Covariates الى النافذة التي تحت Model باستخدام السهم الجانبي ،

- العودة ثانية الى السهم ذات الاتجاه السفلي الموحود في الوسط لاختيار Iteraction
- الكبس على ايقونة Continue للعودة الى مربع الحوار Univariate

الشكل البياني رقم (17.8)

لوحة Univariate : Model لتحليل التباين بمعيارين

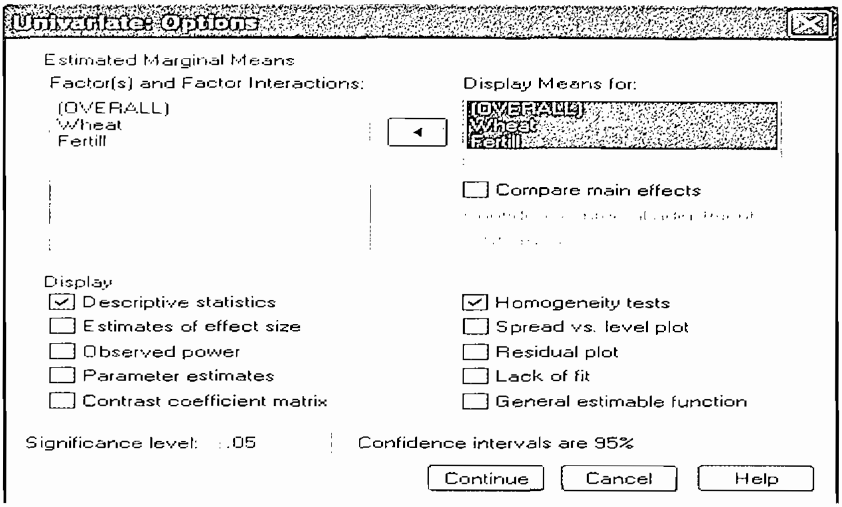


- الكبس على ايقونة Options فتظهر لنا لوحة Univariate Options المبينة في الشكل البياني رقم (18.8)، وفيها يتم الاجراء التالي:

- نقل المتغيرات والتفاعلات المطلوب ايجاد متوسطات لاصناف المتغير التابع، من النافذة التي تحت Factors and Factor Interations الى النافذة التي تحت Display Means for
- تحت Display يتم التاشيرعند Descriptive Statistics للحصول مقاييس المتوسطات والانحراف المعياري، والتاشير كذلك عند Homogeneity Tests لاختبار تجانس تباين اصناف العوامل ،
- الكبس على ايقونة Continue للعودة الى مربع الحوار Univariate

الشكل البياني رقم (18.8)

لوحة Univariate Options لتحليل التباين بمعيارين

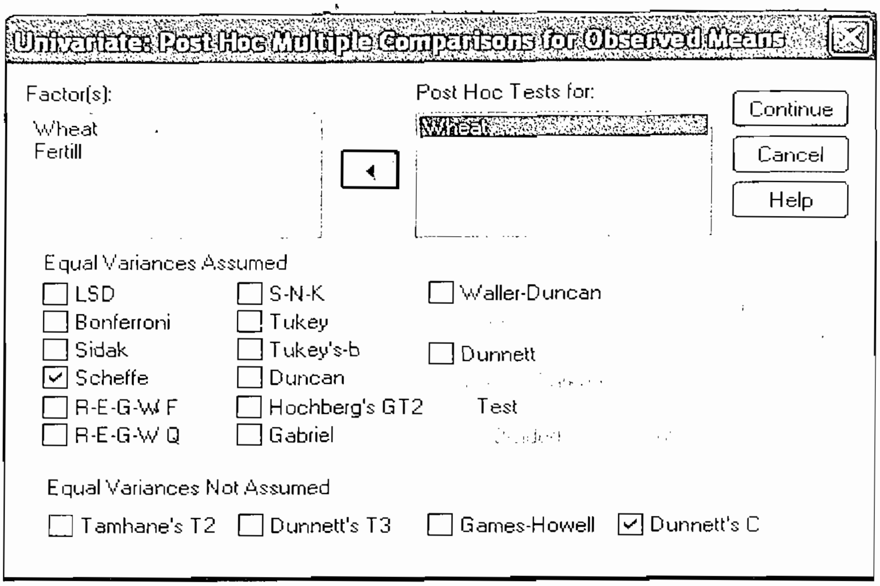


▪ الكبس على ايقونة Post Hoc فتظهر لنا لوحة Post Hoc Multiple Comparisons المبينة في الشكل البياني رقم (19.8)، وفيها يتم الاجراء التالي :

- نقل المتغير المكون من ثلاثة اصناف فاكثر من النافذة التي تحت Factors الى النافذة التي تحت Post Hoc Tests for لاجراء البعدية لاصناف المتغير الذي يتم نقله ،
- التاشير عند Scheffe للمقارنات البعدية لفرضية تساوي تباين الاصناف، تحت Equal Variance Assumed، وعند Dunnett'C لفرضية عدم تساوي تباينات الاصناف، تحت Equal Variance Not Assumed
- الكبس على ايقونة Continue للعودة الى مربع الحوار Univariate من جديد

الشكل البياني رقم (19.8)

لوحة Post Hoc Multiple Comparisons لتحليل التباين بمعيارين



- الكبس على ايقونة Ok لنحصل على ملحق مخرجات التحليل المبين في الجداول رقم (7.8) التالية. ومنها نستدل على وجود فروق ذات دلالة سواء بين اصناف القمح او بين انواع الاسمدة المستخدمة، حيث جاءت قيم f عند اقل من 0.05، ومثل هذه الفروق جاءت واضحة من الجداول التي تضمنتها مخرجات التحليل سواء بين المتوسطات او التباينات وكذلك في اختبار التجانس، وهو ما يتفق مع ما تم الحصول عليه عند حل المثال في اعلاه يدويا لكن من دون تفاعل تفاعل داخلي .

جداول رقم (7.8)

مخرجات تحليل التباين بمعيارين

Two Ways Analysis of Variance

Between-Subjects Factors

		N
Wheat	1.00	3
	2.00	3
	3.00	3
	4.00	3
Fertill	1.00	4
	2.00	4
	3.00	4

Descriptive Statistics

Dependent Variable: Yi

Wheat	Fertill	Mean	Std. Deviation	N
1.00	1.00	9.0000	1.00000	3
	Total	9.0000	1.00000	3
2.00	1.00	7.0000	.	1
	2.00	6.0000	1.41421	2
	Total	6.3333	1.15470	3
3.00	2.00	6.5000	2.12132	2
	3.00	4.0000	.	1
	Total	5.6667	2.08167	3
4.00	3.00	4.3333	.57735	3
	Total	4.3333	.57735	3
Total	1.00	8.5000	1.29099	4
	2.00	6.2500	1.50000	4
	3.00	4.2500	.50000	4
	Total	6.3333	2.10339	12

Levene's Test of Equality of Error Variances(a)

F	df1	df2	Sig.
4.275	5	6	.053

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: Intercept+Wheat+Fertill

Tests of Between-Subjects Effects

Dependent Variable: Yi

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39.500(a)	5	7.900	5.171	.035
Intercept	481.333	1	481.333	315.05	.000
Wheat	3.333	3	1.111	.727	.572
Fertill	4.833	2	2.417	1.582	.281
Error	9.167	6	1.528		
Total	530.000	12			
Corrected Total	48.667	11			

a R Squared = .812 (Adjusted R Squared = .655)

1. Grand Mean

Dependent Variable: Yi

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
6.333	.357	5.460	7.206

2. Wheat

Dependent Variable: Yi

Wheat	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.00	7.500	1.335	4.233	10.767
2.00	5.500	.874	3.361	7.639
3.00	6.000	.874	3.861	8.139
4.00	6.333	1.335	3.067	9.600

3. Fertill

Dependent Variable: Yi

Fertill	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.00	7.833	1.183	4.938	10.729
2.00	6.833	.798	4.881	8.786
3.00	4.333	1.183	1.438	7.229

Wheat: Post Hoc Tests

Multiple Comparisons

	(I) Wheat	(J) Wheat	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	**1.00**	**2.00**	**2.6667**	**1.00922**	**.174**	**-1.1459**	**6.4792**
		3.00	**3.3333**	**1.00922**	**.084**	**-.4792**	**7.1459**

		4.00	4.6667(*)	1.00922	.021	.8541	8.4792
	2.00	1.00	-2.6667	1.00922	.174	-6.4792	1.1459
		3.00	.6667	1.00922	.929	-3.1459	4.4792
		4.00	2.0000	1.00922	.355	-1.8125	5.8125
	3.00	1.00	-3.3333	1.00922	.084	-7.1459	.4792
		2.00	-.6667	1.00922	.929	-4.4792	3.1459
		4.00	1.3333	1.00922	.648	-2.4792	5.1459
	4.00	1.00	-4.6667(*)	1.00922	.021	-8.4792	-.8541
		2.00	-2.0000	1.00922	.355	-5.8125	1.8125
		3.00	-1.3333	1.00922	.648	-5.1459	2.4792
Dunnett C	1.00	2.00	2.6667	.88192		-3.4435	8.7768
		3.00	3.3333	1.33333		-5.9044	12.5710
		4.00	4.6667(*)	.66667		.0478	9.2855
	2.00	1.00	-2.6667	.88192		-8.7768	3.4435
		3.00	.6667	1.37437		-8.8553	10.1887
		4.00	2.0000	.74536		-3.1640	7.1640
	3.00	1.00	-3.3333	1.33333		-12.5710	5.9044
		2.00	-.6667	1.37437		-10.1887	8.8553
		4.00	1.3333	1.24722		-7.3077	9.9744
	4.00	1.00	-4.6667(*)	.66667		-9.2855	-.0478
		2.00	-2.0000	.74536		-7.1640	3.1640
		3.00	-1.3333	1.24722		-9.9744	7.3077

Based on observed means.

* The mean difference is significant at the .05 level.

Homogeneous Subsets

	Wheat	N	Subset	
			1	2
Scheffe(a,b)	4.00	3	4.3333	
	3.00	3	5.6667	5.6667
	2.00	3	6.3333	6.3333
	1.00	3		9.0000
	Sig.		.355	.084

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 1.528.

a Uses Harmonic Mean Sample Size = 3.000.

b Alpha = .05.

الملاحق

ملحق رقم (1.2)

مقطع من التصنيف الدولي للانشطة الاقتصادية ISIC

ويشتمل التصنيف القياسي الدولي لانشطة الاقتصادية والاجتماعية على تسعة اقسام رئيسية هي :

(1) الزراعة والصيد واغابات وصيد الاسماك

(2) الصناعات الاستخراجية (المناجم واستخراج البترول والعاز الطبيعي-صناعات)

(3) الصناعات التحويلية

(4) التشييد والبناء

(5) الكهرباء والماء والغاز والبخار

(6) تجارة الجملة والمفرد والمطاعم والفنادق

(7) النقل والتخزين والمواصلات

(8) التمويل والتامين والعقارات والخدمات المقدمة لقطاع العمل

(9) خدمات المجتمع والخدمات الاجتماعية والشخصية

ولكل من الاقسام الواردة في اعلاه مكونات لحد 4 مسنويات او اكثر طبقا لحاجة كل دولة، فمثلا مستويات قطاع الصناعات التحويلية تاخذ الصيغة التالية :

بيان النشاط	نشاط	فرع	فصل	باب
قسم 3 : الصناعات التحويلية				
صناعة المواد الغذائية والمشروبات والتبغ				31
صناعة المواد الغذائية			311.321	
المجازر وتهيئة وحفظ اللحوم		3111		
ذبح الحيوانات	31111			
ذبح الدواجن	31112			
تهيئة لحوم الحيوانات وصناعة الباسطرمة وما شابهها	31113			
اخرى غير مذكورة	31119			
منتجات الالبان		3112		
يسترة الحليب وتعقيمه وتهيئته في زجاجات	31121			
الزبدة والقشطة	31122			
صناعة اللبن وبساراه	31123			
صماعة الجبن بانواعه	31124			
الايس كريم (مثلجات، جيلاتي ....)	31125			
منتجات الالبان غير المذكورة	31129			
تعليب وحفظ الخضراوات		3113		
تعليب وحفظ الخضراوات	31131			
المربى والشرابت	31132			
تعليب وكبس التمور	31133			
صناعة الدبس	31134			
اخرى غير مذكورة	31139			

ملحق رقم (2.2)

مقطع من التصنيف القياسي الدولي للتعليم *ISCE

الرمز	مستوى التعليم
0	التعليم السابق للمستوى الأول (رياض الأطفال)
1	التعليم في المستوى الأول (التعليم الابتدائي)
2	التعليم في المستوى الثاني (المرحلة الأولى : المتوسطة)
2	التعليم في المستوى الثاني (المرحلة الأولى : الإعدادي)
5	التعليم في المستوى الثالث (المرحلة الأولى : شهادة لتعادل الجامعية)
6	التعليم في المستوى الثالث (المرحلة الأولى : شهادة جامعية أولية)
7	التعليم في المستوى الثالث (المرحلة الثانية : دراسات عليا)
9	التعليم الذي لايمكن تحديد مستواه

*** تقوم منظمة اليونسكو بإصدار ومتابعة هذا الدليل كأداة لتوحيد مفاهيم المعطيات المتعلقة بالتعليم، ويعتمد التصنيف على معيارين أساسيين هما : المستوى الدراسي، و المواد الدراسية المقررة للمستوى .**

ملحق رقم (3.2)

مقطع من التصنيف القياسي الدولي للمهن ISCO

وهو دليل جامع لتفاصيل الأعمال والواجبات المتصلة بالمهن والحرف، وترتيبها في مجموعات وأعطاء رمز لكل مجموعة، وكل مهنة برمز رقمي يعكس العلاقة بينها ويمنع أي تداخل محتمل بين مهنة وأخرى. ويضم الدليل 8 أقسام هي :

القسم 0 / 1 الاختصاصيون والفنيون ومن يرتبط بهم

القسم 2 التشريعيون والرؤساء الإداريون والمدراء

القسم 3 الموظفون التنفيذيون والكتب ومن يرتبط بهم

القسم 4 العاملون في البيع

القسم 5 العاملون في الخدمات

القسم 6 العاملون في الزراعة وتربية الحيوانات والغابات والصيادون وصيادو الأسماك

القسم 7/ 8/ 9 العاملون في الإنتاج ومن يرتبط بهم ومشغلو معدات النقل والتشغيل

القسم / س العاملون الذين لايمكن تصنيفهم

ويضم كل قسم مجموعات فرعية تسمى 'باب' وكل باب وحدات مهنية تسمى 'فصل'، والفصل يضم اختصاصات مهنية تسمى 'مهنة'.

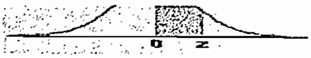
ملحق رقم (1.3)

جدول لارقام عشوائية Random numbers

Hill AB (1977) *A Short Textbook of Medical Statistics.* London: Hodder and Stoughton, 1977:306-7.

	1 2 3 4	5 6 7 8	9 10 11 12	13 14 15 16	17 18 19 20	21 22 23 24	25 26 27 28	29 30 31 32
1	8 0 9 4	2 5 2 5	8 2 4 7	1 3 4 7	7 4 3 3	3 6 2 0	1 3 9 7	2 1 3 4
2	3 6 6 3	2 1 9 8	9 2 1 1	9 0 4 6	2 6 1 8	2 7 5 1	2 8 3 7	1 0 9 5
3	1 3 3 0	6 3 3 1	3 7 5 3	9 8 9 3	8 7 3 8	8 9 1 6	1 5 3 8	8 5 4 3
4	3 5 6 5	0 0 1 8	2 2 4 3	6 4 3 2	4 7 3 8	6 0 9 5	5 2 9 3	1 6 2 0
5	7 8 5 0	5 9 2 5	5 5 8 8	7 3 1 1	2 1 9 2	4 5 4 5	3 5 3 0	5 5 8 9
6	4 4 9 0	5 4 1 7	9 7 2 7	6 1 5 3	5 9 0 1	4 8 7 2	9 9 9 0	9 8 7 7
7	6 5 4 5	9 1 0 4	9 3 1 8	8 8 1 9	7 5 3 7	2 7 8 5	9 3 7 3	2 4 4 6
8	3 6 2 8	5 9 9 5	1 2 1 5	3 7 5 3	9 2 2 3	5 6 5 8	2 9 4 4	2 8 9 9
9	4 6 6 5	4 8 7 0	7 6 5 4	0 6 1 2	5 6 8 3	4 2 5 1	9 1 3 8	1 7 0 9
10	6 4 9 8	7 5 1 9	0 4 7 4	7 8 1 0	8 8 3 2	9 6 8 2	9 9 7 2	4 0 9 0
11	6 7 2 2	9 8 6 3	3 3 6 1	7 8 7 5	4 8 8 3	1 3 1 6	9 6 7 3	8 8 3 4
12	9 7 4 8	5 3 3 2	5 1 1 5	2 7 2 1	0 0 3 3	9 3 0 3	3 7 1 3	4 0 1 2
13	5 6 4 1	1 4 1 7	1 4 1 9	7 4 3 4	8 1 6 5	7 3 6 5	1 2 1 8	5 0 3 9
14	7 4 4 4	9 2 0 0	8 6 4 0	5 6 8 2	4 3 9 8	3 9 0 4	9 1 9 0	3 3 3 8
15	8 2 7 9	3 0 1 9	4 6 7 2	3 7 4 3	3 9 7 9	4 6 8 9	9 0 2 1	6 9 9 0
16	0 1 8 1	7 6 1 7	1 0 2 4	2 3 8 7	2 8 9 1	8 6 7 7	1 6 8 5	2 4 8 2
17	7 3 8 2	9 7 5 9	7 6 5 5	6 8 2 4	5 9 7 7	2 0 0 8	5 5 9 6	9 7 4 0
18	7 8 3 0	4 7 1 4	3 6 9 6	3 9 1 9	1 8 0 4	4 0 4 4	1 0 3 4	2 5 9 7
19	9 8 9 7	4 2 1 6	6 5 2 6	4 5 3 5	8 4 3 0	6 2 7 0	9 8 0 5	0 7 6 8
20	1 2 8 1	2 5 1 6	8 6 5 9	2 3 1 0	5 9 3 9	8 7 0 3	9 9 4 1	0 3 5 3
21	3 3 4 7	4 9 3 7	7 6 3 4	2 5 4 3	6 2 3 9	7 4 5 5	2 0 5 5	7 7 9 5
22	4 5 5 0	8 1 0 3	1 2 5 0	2 3 0 4	1 1 3 8	9 7 8 8	9 1 4 4	4 6 2 8
23	1 3 4 4	3 5 9 7	2 3 8 3	6 9 7 6	8 2 5 1	4 2 0 1	2 0 3 8	6 5 5 2
24	8 3 7 5	5 8 2 3	5 4 8 7	0 4 5 0	3 1 0 8	9 1 5 5	2 7 1 7	7 6 0 1
25	7 7 1 0	9 9 4 3	6 9 7 8	8 2 7 3	9 7 1 4	9 7 0 0	1 6 8 8	2 8 9 9
26	6 3 5 5	6 0 0 5	9 4 4 3	2 2 8 2	1 5 2 4	2 5 1 7	5 8 1 5	0 0 8 1
27	7 3 4 1	2 3 1 2	2 4 2 1	6 7 0 2	9 9 8 4	3 4 5 9	3 0 8 5	4 7 6 2
28	2 2 8 4	0 9 9 6	9 1 0 7	5 5 4 2	7 3 1 9	3 7 8 2	1 0 6 8	9 5 7 4
29	9 5 9 4	7 4 1 5	9 3 6 5	6 0 4 5	1 1 8 3	5 9 1 6	9 5 9 9	1 1 4 3
30	4 6 1 3	8 5 4 9	6 3 6 9	3 2 0 8	5 1 0 9	9 6 9 0	1 1 6 8	6 1 3 3

الملحق رقم (2.3)

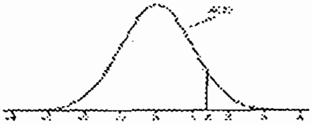


التوزيع الطبيعي الاحتمالي للمساحة الواقعة بين قيم Z والمتوسط

ز	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000
4.0	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000

الملحق رقم (3.3)

التوزيع الطبيعي Z التجميعي الذي يعطي احتمال المتغير العشوائي



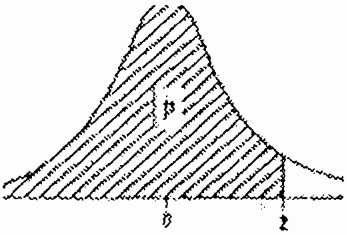
الموزع طبيعيا N(0,1)

Cumulative Standardized Normal Distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999							

ملحق رقم (4.3)

قيم z الجدولية الموزعة طبيعيا $N(0,1)$ عند مستويات معنوية مختلفة



p	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.50	0.000	0.025	0.050	0.075	0.100	0.126	0.151	0.176	0.202	0.228
0.60	0.253	0.279	0.305	0.332	0.358	0.385	0.412	0.440	0.468	0.496
0.70	0.524	0.553	0.583	0.613	0.643	0.674	0.706	0.739	0.772	0.806
0.80	0.842	0.878	0.915	0.954	0.994	1.036	1.080	1.126	1.175	1.227
0.90	1.282	1.341	1.405	1.476	1.555					

p	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
0.95	1.645	1.655	1.665	1.675	1.685	1.695	1.706	1.717	1.728	1.739
0.96	1.751	1.762	1.774	1.787	1.799	1.812	1.825	1.838	1.852	1.866
0.97	1.881	1.896	1.911	1.927	1.943	1.960	1.977	1.995	2.014	2.033
0.98	2.054	2.075	2.097	2.120	2.144	2.170	2.197	2.226	2.257	2.290
0.99	2.326	2.366	2.409	2.457	2.512	2.576	2.652	2.748	2.878	3.090

الملحق رقم (5.3)

قيم t الجدولية عند مستويات معنوية مختلفة ودرجات الحرية v

t Distribution: Critical Values of t

df	90%	95%	97.5%	99%	99.5%	99.9%
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.183	4.541	5.841	10.215
4	1.533	2.132	2.777	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.708	5.208
7	1.415	1.895	2.365	2.998	3.500	4.785
8	1.397	1.860	2.306	2.897	3.355	4.501
9	1.383	1.833	2.262	2.822	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.625	2.977	3.787
15	1.341	1.753	2.132	2.603	2.947	3.733
16	1.337	1.746	2.120	2.584	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.611
19	1.328	1.729	2.093	2.540	2.861	3.580
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.788	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.705	3.307
80	1.292	1.664	1.990	2.374	2.639	3.195
∞	1.282	1.645	1.960	2.326	2.576	3.090

الملحق رقم (6.3)

قيم مربع كاي χ^2 عند عدد من مستويات معنوية ودرجات الحرية v

df \ X	0.999	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005	0.001
1	1.6 E-6	3.93E-5	0.00016	0.00098	0.00393	3.84	5.02	6.63	7.88	10.83
2	0.002	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.6	13.82
3	0.02	0.07	0.11	0.22	0.35	7.81	9.35	11.34	12.84	16.27
4	0.09	0.21	0.30	0.48	0.71	9.49	11.14	13.28	14.86	18.47
5	0.21	0.41	0.55	0.83	1.15	11.07	12.83	15.09	16.75	20.52
6	0.38	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55	22.46
7	0.60	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28	24.32
8	0.86	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.95	26.12
9	1.15	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59	27.88
10	1.48	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19	29.59
11	1.83	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76	31.26
12	2.21	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.3	32.91
13	2.62	3.57	4.11	5.01	5.89	22.36	24.74	27.69	29.82	34.53
14	3.04	4.07	4.66	5.63	6.57	23.68	26.12	29.14	31.32	36.12
15	3.48	4.60	5.23	6.26	7.26	25.00	27.49	30.58	32.80	37.70
16	3.94	5.14	5.81	6.91	7.96	26.30	28.85	32.00	34.27	39.25
17	4.42	5.70	6.41	7.56	8.67	27.59	30.19	33.41	35.72	40.79
18	4.90	6.26	7.01	8.23	9.39	28.87	31.53	34.81	37.16	42.31
19	5.41	6.84	7.63	8.91	10.12	30.14	32.85	36.19	38.58	43.82
20	5.92	7.43	8.26	9.59	10.85	31.41	34.17	37.57	40.00	45.31
21	6.45	8.03	8.90	10.28	11.59	32.67	35.48	38.93	41.4	46.80
22	6.98	8.64	9.54	10.98	12.34	33.92	36.78	40.29	42.80	48.27
23	7.53	9.26	10.20	11.69	13.09	35.17	38.08	41.64	44.18	49.73
24	8.08	9.89	10.86	12.40	13.85	36.42	39.36	42.98	45.56	51.18
25	8.65	10.52	11.52	13.12	14.61	37.65	40.65	44.31	46.93	52.62
26	9.22	11.16	12.20	13.84	15.38	38.89	41.92	45.64	48.29	54.05
27	9.80	11.81	12.88	14.57	16.15	40.11	43.19	46.96	49.64	55.48
28	10.39	12.46	13.56	15.31	16.93	41.34	44.46	48.28	50.99	56.89
29	10.99	13.12	14.26	16.05	17.71	42.56	45.72	49.59	52.34	58.30
30	11.59	13.79	14.95	16.79	18.49	43.77	46.98	50.89	53.67	59.70
31	12.20	14.46	15.66	17.54	19.28	44.99	48.23	52.19	55.00	61.10
32	12.81	15.13	16.36	18.29	20.07	46.19	49.48	53.49	56.33	62.49
33	13.43	15.82	17.07	19.05	20.87	47.40	50.73	54.78	57.65	63.87
34	14.06	16.50	17.79	19.81	21.66	48.60	51.97	56.06	58.96	65.25
35	14.69	17.19	18.51	20.57	22.47	49.80	53.20	57.34	60.27	66.62
36	15.32	17.89	19.23	21.34	23.27	51.00	54.44	58.62	61.58	67.99
37	15.97	18.59	19.96	22.11	24.07	52.19	55.67	59.89	62.88	69.35
38	16.61	19.29	20.69	22.88	24.88	53.38	56.90	61.16	64.18	70.70
39	17.26	20.00	21.43	23.65	25.70	54.57	58.12	62.43	65.48	72.05

ملحق رقم (7.3)

قيم f الجدولية عند عدد من مستويات المعنوية ودرات الحرية V_1 و V_2

F Distribution: Critical Values of F (5% significan

v_2 \ v_1	25	30	35	40	50	60	75	100	150	200
1	249.26	250.10	250.69	251.14	251.77	252.20	252.62	253.04	253.46	253.68
2	19.46	19.46	19.47	19.47	19.48	19.48	19.48	19.49	19.49	19.49
3	8.63	8.62	8.60	8.59	8.58	8.57	8.56	8.55	8.54	8.54
4	5.77	5.75	5.73	5.72	5.70	5.69	5.68	5.66	5.65	5.65
5	4.52	4.50	4.48	4.46	4.44	4.43	4.42	4.41	4.39	4.39
6	3.83	3.81	3.79	3.77	3.75	3.74	3.73	3.71	3.70	3.69
7	3.40	3.38	3.36	3.34	3.32	3.30	3.29	3.27	3.26	3.25
8	3.11	3.08	3.06	3.04	3.02	3.01	2.99	2.97	2.96	2.95
9	2.89	2.86	2.84	2.83	2.80	2.79	2.77	2.76	2.74	2.73
10	2.73	2.70	2.68	2.66	2.64	2.62	2.60	2.59	2.57	2.56
11	2.60	2.57	2.55	2.53	2.51	2.49	2.47	2.46	2.44	2.43
12	2.50	2.47	2.44	2.43	2.40	2.38	2.37	2.35	2.33	2.32
13	2.41	2.38	2.36	2.34	2.31	2.30	2.28	2.26	2.24	2.23
14	2.34	2.31	2.28	2.27	2.24	2.22	2.21	2.19	2.17	2.16
15	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12	2.10	2.10
16	2.23	2.19	2.17	2.15	2.12	2.11	2.09	2.07	2.05	2.04
17	2.18	2.15	2.12	2.10	2.08	2.06	2.04	2.02	2.00	1.99
18	2.14	2.11	2.08	2.06	2.04	2.02	2.00	1.98	1.96	1.95
19	2.11	2.07	2.05	2.03	2.00	1.98	1.96	1.94	1.92	1.91
20	2.07	2.04	2.01	1.99	1.97	1.95	1.93	1.91	1.89	1.88
21	2.05	2.01	1.98	1.96	1.94	1.92	1.90	1.88	1.86	1.84
22	2.02	1.98	1.96	1.94	1.91	1.89	1.87	1.85	1.83	1.82
23	2.00	1.96	1.93	1.91	1.88	1.86	1.84	1.82	1.80	1.79
24	1.97	1.94	1.91	1.89	1.86	1.84	1.82	1.80	1.78	1.77
25	1.96	1.92	1.89	1.87	1.84	1.82	1.80	1.78	1.76	1.75
26	1.94	1.90	1.87	1.85	1.82	1.80	1.78	1.76	1.74	1.73
27	1.92	1.88	1.86	1.84	1.81	1.79	1.76	1.74	1.72	1.71
28	1.91	1.87	1.84	1.82	1.79	1.77	1.75	1.73	1.70	1.69
29	1.89	1.85	1.83	1.81	1.77	1.75	1.73	1.71	1.69	1.67
30	1.88	1.84	1.81	1.79	1.76	1.74	1.72	1.70	1.67	1.66
35	1.82	1.79	1.76	1.74	1.70	1.68	1.66	1.63	1.61	1.60
40	1.78	1.74	1.72	1.69	1.66	1.64	1.61	1.59	1.56	1.55
50	1.73	1.69	1.66	1.63	1.60	1.58	1.55	1.52	1.50	1.48
60	1.69	1.65	1.62	1.59	1.56	1.53	1.51	1.48	1.45	1.44
70	1.66	1.62	1.59	1.57	1.53	1.50	1.48	1.45	1.42	1.40
80	1.64	1.60	1.57	1.54	1.51	1.48	1.45	1.43	1.39	1.38
90	1.63	1.59	1.55	1.53	1.49	1.46	1.44	1.41	1.38	1.36
100	1.62	1.57	1.54	1.52	1.48	1.45	1.42	1.39	1.36	1.34
120	1.60	1.55	1.52	1.50	1.46	1.43	1.40	1.37	1.33	1.32
150	1.58	1.54	1.50	1.48	1.44	1.41	1.38	1.34	1.31	1.29
200	1.56	1.52	1.48	1.46	1.41	1.39	1.35	1.32	1.28	1.26
250	1.55	1.50	1.47	1.44	1.40	1.37	1.34	1.31	1.27	1.25
300	1.54	1.50	1.46	1.43	1.39	1.36	1.33	1.30	1.26	1.23
400	1.53	1.49	1.45	1.42	1.38	1.35	1.32	1.28	1.24	1.22
500	1.53	1.48	1.45	1.42	1.38	1.35	1.31	1.28	1.23	1.21
600	1.52	1.48	1.44	1.41	1.37	1.34	1.31	1.27	1.23	1.20
750	1.52	1.47	1.44	1.41	1.37	1.34	1.30	1.26	1.22	1.20
1000	1.52	1.47	1.43	1.41	1.36	1.33	1.30	1.26	1.22	1.19

ملحق رقم (1.6)

القيم الجدولية لاختبار دارين- واتسون عند $\alpha=0.05$ و $\alpha=0.01$ وفقاً لعدد المشاهدات وعدد المتغيرات المنقلة k

For $\alpha=0.05$

	$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

تابع ملحق رقم (1.6)

For α=0.01

	k = 1		k = 2		k = 3		k = 4		k = 5	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

المراجع

مراجع باللغة العربية

1. البلداوي عبدالحميد، 2004، **اساليب البحث العلمي والتحليل الاحصائي باستخدام برنامج SPSS**، دار الشروق للنشر والتوزيع – عمان .
2. البلداوي عبدالحميد، 1997، **الإحصاء للعلوم الإدارية والتطبيقية**، دار الشروق للنشر والتوزيع / عمان – الأردن.
3. الزعبي محمد بلال والطلافحة عباس، 2003، **النظام الاحصائي SPSS**، دار وائل للنشر، عمان-الاردن .
4. غرايبة فوزي واخرين، 2002، **اساليب البحث العلمي في العلوم الاجتماعية والانسانية**، الطبعة الثالثة، دار وائل للنشر، عمان – الاردن.

مراجع باللغة الانكليزية

1. Busbas, D.S. (1979) "A multidimensional Scaling Approach to the Determination of Preferences for Transportation projects " , Ph.D. thesis, indiana University U.S.A.
2. Cochron William G. , 1980 , Sampling Techniques, New York, Jon Wiley.
3. Deming W.Edwards, 1980, Sample Design in Business Research, Wiley, New York.
4. Draper N R and Smith H , 1980 , Applied Regression Analysis , 3rd Ed., John Wiley and Sons inc., London .
5. Daling , J.R. and Tomura, 1970 , Use of Orthogonal Factors for selection of Variables in a Regression Equation An illustration ,Journal of Applied Statistics. 19.
6. Draper & Smith, 1990, Applied Regression Analysis, John Wiley and son Inc, London .
7. Fishbein, M. (1967). " Reading in attitude theory and measurement ' John Wily and Sons Inc.
8. Fishbein , M. and Ajzen,(1975). ' Beliefs, attitudes, intention and behavior : an introduction to theory and research '. Addison-Wesley, Reading, Mass.
9. Hartgen , D.T. (1973). ' The influence of attitudinal and Situational Variables on Urban mode choice ", Ph.D. thesis, Urban and Reginal Planning, northwestern University.

10. Jeffers, J.P. An Introduction to system Analysis : with ecological applications, William Clowes and sons LTD, London , 1978 .

11.Kendall M, 1981, Multivariate Analysis, 2nd Ed., Charls Greffin and Company Ltd., London

12.Koutsoyiannis, A. (1977).'Theory of Econometrics' , second edition, The Macmillan Press LTD., New York .

13..Morrison, D.F. , Multivariate Statistical Methods, Mc. Graw-Hill, New York, 1967

14.Torgerson , W.S. (1958) , ' Theory and methods of scaling ' John Wiely and Sons, Inc. London .

15.W.J. Krzanowski, Principles of Multivariate Analysis, Oxford University Press, 1988 .

16. Zorkovich S S, 1981, Presentation of Surveyes Proccedings of the 3rd Session, Bulletin of the International Statistical Institute, Buenos Aires, Book 1.

مجلات علمية

1. Brand D. (1976). ' Approaches to Travel Behavior Research' Transportation Research , 567, pp 12-33 .

2. Burbett,p.(1973). ' The Dimensions of alternative in spatial choice processes ', Geographicl Analysis, Vol. 3, pp 181-204 .

3. Hocking, R.R., The Analysis and Selection of Variables in Linear Regression Biometrics, 32, PP. 1-49, 1976

4. Kansky, K.J. (1967). 'Travel pattern of Urban residents ', Transportation Science Vol. I, PP 261-258 .

5. paine, F.T.et.al (1969).'Consumer attitudes toward auto versus public Transport alternatives, Journal of Applied Psychology , Vol.6,PP 472-480

"المؤلف في سطور"

الدكتورعبدالحميد عبدالمجيد البلداوي

FSS ; MCIT ; MASA

beldawin@yahoo.ca

- مواليد بغداد – العراق في 5/ 9/ 1945 .
- حاصل على الدكتوراه والماجستير (بريطانيا) والبكالوريوس من العراق في اختصاص الاحصاء التطبيقي .
- عمل باحث وخبير ومدير باحثين في مجال التخطيط والاحصاء لمدة 26 سنة في العراق ودولة الامارات .
- عمل استاذ مشارك لمدة 15 سنة في جامعات: عراقية– اردنية– ليبية
- ساهم بدراسات لاغراض الامم المتحدة ومؤسسات احصائية عربية. وفي العديد من المؤتمرات الدولية والعربية منها اليونسكو- الاسكوا- بريطانيا- الجامعة العربية- المنظمة العربية للتنمية الادارية – منظمة مؤتمرالاقطار الاسلامية وغيرها، وفي مواسم علمية وثقافية في جامعات ومؤسسات عربية عديدة، وفي اقامة العديد من الدورات التدريبية في مجال: اتخاذ القرارات باستخدام النماذج الاحصائية – تصميم العينات وتطبيقها– التحليل باستخدام برنامج SPSS – بناء الارقام القياسية واستخداماتها .
- نشر 20 بحثا،
- تأليف الكتب التالية :

1. ألاساليب التطبيقية لتحليل واعداد البحوث العلمية " مع حالات دراسية باستخدام برنامج SPSS"، 2008، دار الشروق للنشر والتوزيع، عمان .
2. الإحصاء للعلوم الإدارية والتطبيقية، 1997، دار الشروق للنشر والتوزيع، عمان – الأردن
3. اساليب البحث العلمي والتحليل الاحصائي باستخدام برنامج SPSS، 2004 ،دار الشروق للنشر والتوزيع – عمان .
4. الاساليب الاحصائية التطبيقية، دار الشروق للنشر – عمان، 2004 .
5. تطبيقات الحاسوب في العمليات الادارية والمالية "مشترك"، 2004 ،دار الشروق للنشر والتوزيع – عمان،
6. الطرق الاحصائية التطبيقية للمعاينة، 1995 ،جامعة السابع من ابريل، ليبيا .
7. إدارة الجودة الشاملة والمعولية (الموثوقية)، التقنيات الحدثية في التطبيق والاستدامة، 2007 "مشترك" دار الشروق للنشر والتوزيع، عمان.
8. الاساليب الكمية في ادارة الاعمال، 2008 "مشترك"، دار وائل للنشر والتوزيع، عمان
9. الإحصاء للباحثين والمخططين "مشترك" / معهد التخطيط القومي – وزارة التخطيط / مطبعة الجاحظ - بغداد/ 1985
10. التطور النوعي والمالي لقطاع النقل في العراق، 1971 ،وزارة التخطيط، بغداد .

الأساليب التطبيقية

لتحليل وإعداد البحوث العلمية

مع حالات دراسية باستخدام برنامج SPSS

ان هذا الكتاب هو واحد من حصيلة المزاوجة بين الخبرة الأكاديمية والخبرة العملية في مجالات البحوث والتخطيط ومن بين الحصيلة أيضاً اليقين من أن التشعب والتعقيد لا ييسران للباحث تعبيد طريق ، وبذلك حاولنا في هذا الكتاب التبسيط والتيسير وبتسلسل منطقي نحو المتطلبات العملية في اعداد الدراسات والبحوث . معززين كل ما يذكر في الكتاب بحالات دراسية من اعداد المؤلف بأمل ان تكون وافية قدر الامكان ، مبتعدين عن الخوض في مفاصل نظرية غير ضرورية وإيلاء الاهتمام والتركيز على المواضيع التطبيقية .

ISBN 9957-00-338-0

9789957003388

عمان-الأردن / رام الله - فلسطين